REMARKS/ARGUMENTS

Claims 1-3, 9-12, 17-19, 38-40, 43-46, 49-52, and 55-64 are pending in the application. New claim 65 has been added. Support for the newly-added claim 65 can be found in the specification, for example, on page 17, lines 13-20, and page 17, line 28 through page 18, line 2. Also, support for fragments or truncated proteins can be found in Example 4 (specification pp. 65-66), in which both the full-length endotoxin encoded by SEQ ID NO:1 and a truncated protein encoded by SEQ ID NO:15 were assayed for pesticidal activity against southern corn rootworm. The nucleotide sequence of SEQ ID NO:15 is a truncation of SEQ ID NO:1 which shares about 55% sequence identity with SEQ ID NO:1. Further support is provided in Example 6 (specification pp. 67-69), in which several truncated proteins were assayed and shown to have pesticidal activity against Colorado potato beetle (see Table 1, p. 68). No new matter has been added by way of amendment. Reexamination and reconsideration of the claims are respectfully requested in view of the discussion herein.

The Rejection of Claims Under 35 U.S.C. §112, First Paragraph, Should Be Withdrawn

The Office Action (10/21/2004, page 2, #4) has rejected claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 under 35 U.S.C. §112, first paragraph, as lacking enablement. While the Office Action states (10/21/2004, page 3, first paragraph) that this "rejection is modified from the rejection set forth in the Office Action mailed 3 December 2003," the articulated bases for the rejection are essentially the same as those set forth previously. Applicants respectfully disagree with this rejection and the articulated bases for this rejection, and further note that this issue was on appeal in this case before prosecution was reopened and was discussed thoroughly in the Appeal Brief.

The rejected claims contain limitations that require the nucleotide sequence of the claims to share a specified percent of sequence identity to SEQ ID NO:1 and thus are referred to herein as "sequence identity claims." The crux of the disagreement about the rejections of the sequence identity claims is that Applicants believe it is unreasonable for the Office Actions to seek to limit the claims to only the exact exemplary sequences disclosed in the specification, because as is well known to those of skill in the art, it is a relatively simple task to modify a nucleotide

sequence and corresponding amino acid sequence at a few positions to generate a protein that retains the full activity of the original—that is, it is easy for one of skill in the art to essentially duplicate the claimed invention using a slightly modified sequence. However, such a sequence would fall outside the literal scope of a claim that was drawn only to the original sequence. Thus, if the claims are limited to the exact exemplary sequence disclosed as suggested in the Office Action (10/21/2004, page 2, #4), Applicants will in fact have taught the public how to make and use the invention without obtaining any meaningful protection which provides the incentive to invention contemplated in the Constitution (Article I, Section 8, clause 8, which gives Congress the power "[t]o promote the Progress of Science and useful Arts, by securing for limited times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.") Accordingly, Applicants respectfully submit that sequence identity claims should be viewed favorably and that the pending claims should be allowed.

As an initial matter, Applicants note that the enablement rejection in the Office Action of 10/21/2004 only refers to the claim limitation that requires nucleotide sequences to have at least 90% sequence identity to SEQ ID NO:1, but the rejected claims include claims specifying at least 93%, 94%, and 95% sequence identity (*i.e.*, claims 55, 58, and 63 (93%), claims 56, 59, and 64 (94%), and claims 38, 43, and 49 (95%)). While Applicants believe that claims would meet the patentability requirements even if they specified percentages of sequence identity well below 90%, Applicants have nevertheless sought to advance prosecution by proposing these claims that require higher degrees of sequence identity to the exemplary disclosed sequences, and by proposing claims with a range of specified sequence identity that is higher than 90%. Unfortunately, the claims with limitations specifying higher than 90% sequence identity have been largely ignored in the Office Actions and were only commented on briefly in the Advisory Action of 2/18/2004. The latest Office Action of 10/21/2004 again simply ignores Applicants' arguments that these claims are enabled. Applicants continue to seek reasonable protection for their invention and respectfully request that each limitation of the pending claims be reconsidered and reexamined on its own merits.

The Office Action (10/21/2004, page 2, #4) rejected claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 under 35 U.S.C. §112, first paragraph, because:

> the specification, while being enabling for nucleic acids encoding SEQ ID NO:2 and 10, expression cassettes comprising the nucleic acids, plants and seeds comprising a construct comprising the nucleic acid and a method of using it to impact a plant pest, does not reasonably provide enablement for any nucleic acid that has 90% identity to SEQ ID NO:1, expression cassettes comprising the nucleic acid, plants and seeds comprising a construct comprising the nucleic acid and a method of using it to impact a plant pest.

Applicants respectfully traverse this rejection and submit that the Examiner is applying an extraordinarily high standard of enablement to the present claims, a standard that is not properly based on case law or on statute.

Support is provided for the limitations of the claims

First, in contrast to the conclusion reached in the Office Action, support is provided for both the sequence identity limitations of the claims (*e.g.*, limitations specifying that said nucleotide sequence has at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1) and the functional limitations of the claims (*i.e.,* that the encoded polypeptide is pesticidal for at least one pest belonging to the order Coleoptera). That is, guidance is provided as to what sequence alterations may be made and still provide a pesticidal polypeptide encompassed by the claim. As discussed further below, endotoxin genes are well known in the art. Applicants have provided the exemplary nucleotide sequence of SEQ ID NO:1 and the exemplary amino acid sequence of SEQ ID NO:2. Indeed, as quoted above, the Office Action (10/21/2004, page 14, #9) stated that the specification was enabling for nucleic acids encoding SEQ ID NO:2 and indicated that claims drawn to the exemplary disclosed sequences (*i.e.*, claims 39, 40, 44, 45, 50, and 51) would be allowable if rewritten in independent form. The claimed sequences of the invention vary from the exemplary disclosed sequences by structural parameters (*i.e.*, percent sequence identity to SEQ ID NO:1; encoding the amino acid sequence set forth in SEQ ID NO:2). Guidance for determining percent identity of sequences is provided in the specification on pages 33 through 38. Thus, support is provided to enable one of skill in the art to make and use a nucleic acid and/or nucleotide construct meeting the sequence identity limitation(s) of the claims.

Support is also provided for the functional limitations of the claims (*i.e.*, that the encoded polypeptide is pesticidal for at least one pest belonging to the order Coleoptera). The independent sequence identity claims (*i.e.*, claims 1, 9, and 17) specify that the nucleotide sequence encodes a polypeptide which is pesticidal for at least one pest belonging to the order Coleoptera; therefore, these claims (and the claims dependent on them) encompass functional variants. Guidance is provided regarding alterations that allow the sequence to retain the specified pesticidal activity (see, *e.g.*, p. 18 (providing guidance regarding conservative substitutions of amino acids) and pp. 19-20 (discussing the activity of variants)). While not every nucleic acid or nucleotide construct that meets the sequence identity limitation(s) of the claims will necessarily meet the functional limitation(s) of the claims, one of skill is readily able to make and use nucleic acids and nucleotide constructs meeting these limitations and thus the claims are enabled. Producing a nucleic acid and/or nucleotide construct that satisfies the functional limitation(s) of the claims is within the skill of one in the art because methods for assaying the pesticidal activity of proteins are routine in the art and because such methods are also described and demonstrated in the specification, for example, on pages 8 and 29 and in the experimental section in working examples such as Example 4 (pp. 65-66), Example 6 (p. 67), and Example 7 (p. 69). Pages 8 and 29 provide support for the limitation requiring that the encoded polypeptide have pesticidal activity, while the working examples teach methods for assaying pesticidal activity of proteins and demonstrate results obtained using these assays. In this manner, Applicants have provided ample guidance regarding the limitations of the claims, including the sequence identity limitations and functional limitations of the claims, and therefore the claims are enabled.

B. thuringiensis δ-endotoxins are well-known in the art, and further support and guidance is provided by working examples

The Office Action (10/21/2004, page 4, third paragraph) states that

> The instant specification fails to provide guidance for which amino acids of SEQ ID NO:2 can be altered and to which other amino acids, and which amino acids must not be changed, to maintain Cry8 activity of the encoded protein. The specification also fails to provide guidance for which amino

> acids can be deleted and which regions of the protein can tolerate insertions and still produce a functional enzyme.

Applicants respectfully disagree with this conclusion. As discussed extensively in the specification (*e.g.*, pp. 3, 7, 11-12, 15, 24-25), the disclosed exemplary sequence of SEQ ID NO:2 is a *Bacillus thuringiensis* Cry-8-like δ-endotoxin. The *B. thuringiensis* δ-endotoxins are an extremely well-characterized group of proteins. As discussed in the specification at pp. 24-25:

> Many of the δ-endotoxins are related to various degrees by similarities in their amino acid sequences and tertiary structure, and means for obtaining the crystal structures of *B. thuringiensis* endotoxins are well known. Exemplary high-resolution crystal structure solution of both the Cry3A and Cry3B polypeptides are available in the literature. The inventors of the present invention used the **solved structure of the *Cry3A* gene** (Li *et al.* (1991) *Nature* 353:815-821) to produce a homology model of the Cry8 δ-endotoxin disclosed and claimed herein as SEQ ID NO:2 to gain insight into the relationship between structure and function of the endotoxin, and to design the recombinantly engineered proteins disclosed and claimed herein. A combined consideration of the **published structural analyses** of *B. thuringiensis* endotoxins and the reported function associated with particular structures, motifs, and the like indicates that specific regions of the endotoxin are correlated with particular functions and discrete steps of the mode of action of the protein. For example, **δ-endotoxins isolated from *B. thuringiensis* are generally described as comprising three domains, a seven-helix bundle that is involved in pore formation, a three-sheet domain that has been implicated in receptor binding, and a beta-sandwich motif** (Li *et al.* (1991) *Nature,* 305: 815-821).

As discussed in more detail in the specification (see, *e.g.*, p. 25), the inventors made use of this knowledge in the art to design specific mutations in the Cry8-like proteins to enhance their pesticidal activity. This strategy was successful in creating altered endotoxins with increased toxicity, as demonstrated by the data presented in working Example 6.

The Office Action (10/21/2004, page 7, first paragraph) dismisses the above-quoted description of the knowledge in the art and Applicants' arguments regarding the knowledge in the art, stating only that:

> Li et al only provide guidance for making truncations and insertion of chymotrypsin cleavage sites; Li et al do not provide guidance for making 362 amino acid substitutions in a 1206 amino acid protein.

Unfortunately, this summarization of the Li reference is incomplete and would be misleading to one unfamiliar with the Li reference. As discussed in the specification (as quoted above), Li teaches the tertiary structure of an exemplary Cry endotoxin. For convenience, a copy of the Li reference is attached hereto as Appendix A. Moreover, as demonstrated by these working examples in the specification, those of skill in the art (*i.e.*, the inventors) were able, in view of the extensive knowledge in the art about *B. thuringiensis* δ-endotoxin structure and function, to modify the novel exemplary sequences disclosed in the specification to provide variant endotoxins with enhanced pesticidal activity. Those of skill in the art, equipped with the novel exemplary sequences disclosed in the specification, would be readily able to further modify those sequences to provide a nucleic acid or nucleotide construct of the invention.

In addition to the tertiary structure of an exemplary Cry endotoxin taught by the Li reference, those of skill in the art are aware of conserved regions of the Cry endotoxins. Provided herewith as Appendix B is an alignment of exemplary SEQ ID NO:2 with the Pfam consensus domains for endotoxins, which are referred to as "Endotoxin N," "Endotoxin M," and "Endotoxin C" (Pfam accession numbers PF03945, PF00555, and PF03944, respectively, descriptions of which are also attached as Appendix B). The Pfam database provides a curated collection of well-characterized protein family domains with high quality alignments. It is well known in the art that regions of sequence homology with known functional domains may be used to determine protein function and to identify what regions of a protein are particularly conserved (and therefore less likely to tolerate mutations) as well as what regions of a protein are less conserved (and therefore more likely to tolerate mutations). Accordingly, Applicants submit that the novel exemplary sequences disclosed in the specification, combined with the knowledge of one familiar with the art, provide adequate guidance as to which amino acids of SEQ ID NO:2 may be modified while allowing the protein to retain pesticidal activity.

As discussed previously, the data and working examples provided in the specification also demonstrate the enablement of the claimed invention by showing that sequences of the

invention that share a relatively low percent identity to the exemplary sequence of SEQ ID NO:1 encode polypeptides that have pesticidal activity against several Coleopteran pests. In Example 4 (specification pp. 65-66), both the full-length endotoxin encoded by SEQ ID NO:1 and a truncated protein encoded by SEQ ID NO:15 were assayed for pesticidal activity against southern corn rootworm. The nucleotide sequence of SEQ ID NO:15 is a truncation of SEQ ID NO:1 which shares about 55% sequence identity with SEQ ID NO:1. In Example 6 (specification pp. 67-69), several truncated proteins were assayed and shown to have pesticidal activity against Colorado potato beetle (see Table 1, p. 68). These truncated proteins included those encoded by SEQ ID NO:15 and SEQ ID NO:19, which share about 55% and 51% identity, respectively, to the exemplary nucleic acid sequence set forth in SEQ ID NO:1 (alignments performed using BLAST with default parameters). The Office Action (page 7, third paragraph) dismisses Applicants' arguments that the truncated proteins provide support as argued, stating that "the local match similarity between SEQ ID NO:15 and the first half of SEQ ID NO:1 is 100%." However, this does not negate the fact that SEQ ID NO:15 is structurally very different from SEQ ID NO:1 and yet provides a protein having pesticidal activity against Coleopteran pests; accordingly, in addition to the knowledge in the art regarding the structure of endotoxins, Applicants have provided guidance in the form of truncated proteins that lack significant portions of the exemplary sequence of SEQ ID NO:1 yet still provide the desired function.

As briefly discussed above, Example 6 also provides assay data for a mutated sequence, NGSR1218-1. This NGSR1218-1 mutant includes the amino acid sequence "NGSR" inserted between amino acids 164 and 165 of the truncated endotoxin of SEQ ID NO:16. The nucleotide sequence encoding this mutant (SEQ ID NO:11) shares about 56% sequence identity with the exemplary nucleotide sequence of SEQ ID NO:1, yet as documented by the data provided in Example 6, both proteins have pesticidal activity. In addition to this data, the specification also provides an exemplary maize-optimized sequence (SEQ ID NO:9) which encodes the same pesticidal polypeptide as SEQ ID NO:15 but shares less than 69% sequence identity with it. Thus, the specification is replete with working examples of sequences that share a relatively low percentage of identity with SEQ ID NO:1 and which encode polypeptides having pesticidal activity. In fact, the percentage of sequence identity shared by the exemplary SEQ ID NO:1 and

these sequences in the working examples is much lower than the "at least 90%" of the broadest sequence identity claims.

The Office Action (10/21/2004, page 8, first paragraph) again dismisses the working examples provided by Applicants, concluding that the specification teaches only "a fragment and a single 4 amino acid insertion" and that "Applicant has not provided guidance for making up to 362 amino acid substitutions in a 1206 amino acid protein." Applicants respectfully disagree with this conclusion. Applicants have provided percent identity variants that include both fragments and amino acid changes to the exemplary wildtype sequences of SEQ ID NO:1 and the encoded SEQ ID NO:2 and thus have taught representative species of the genus of sequences having a particular structural relationship to the exemplary wildtype sequences. Further, as discussed above, there is extensive knowledge in the art about the structure and function of endotoxins and therefore one of skill in the art equipped with the novel sequences disclosed in the present specification would readily be able to identify which portions of the exemplary disclosed sequence are more conserved and which portions would be likely to tolerate mutations. Accordingly, sufficient guidance has been provided to enable the claimed invention.

The amount of experimentation required to make and use the subject matter of the claims is not undue

The Office Action (10/21/2004, page 8, fourth paragraph) concludes that "undue trial and error experimentation would be required to make and assay vast numbers of nucleic acids in order to find any that fell within the scope of the claims." The Office Action states in support of this conclusion that "each assay [as detailed in Examples 4, 6, and 7] requires up to two weeks and large quantities of materials." Applicants believe that this degree of experimentation is not undue. Support for the assertion that such experimentation is not undue is provided by the descriptions detailed in the cited working examples and also by reports in the art of similar experimentation, such as the research reported in Wu *et al.* (2000) *FEBS Letters* 473: 227-232, previously cited in the IDS filed April 15, 2002 and entitled, "Enhanced toxicity of *Bacillus thuringiensis* Cry3A δ-endotoxin in coleopterans by mutagenesis in a receptor binding loop." Because experiments requiring essentially the same amount of work as that required to practice

the claimed invention are regularly reported in the art, Applicants respectfully submit that the amount of experimentation required to practice the claimed invention is not undue.

As discussed previously, the Federal Circuit has repeatedly stated that enablement is not precluded by the necessity for some experimentation, so long as the experimentation needed to practice the invention is not undue, and that a considerable amount of experimentation is permissible if it is merely routine or if the specification provides a reasonable amount of guidance as to how the experimentation should proceed. *Id.* *In re Wands*, 858 F.2d 731, 8 USPQ2d 1400 (Fed Cir 1988). In the instant case, the quantity of experimentation required to practice independent claim 1 amounts to two steps: (1) generating a nucleic acid comprising a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO:1; and (2) assaying the encoded polypeptide for functional activity. Such assays, while known in the art, have further been presented in the specification. One of skill in the art would appreciate that both of these steps are within the skill of those in the art and that this degree of experimentation is not considered undue. In support of this assertion, Applicants are providing herewith as Appendix C a Rule 132 declaration from inventor Dr. André Abad.

Similarly, the amount of experimentation needed to practice the other sequence identity claims is not undue. For example, independent claim 9 recites a transformed plant comprising a nucleotide construct that has a nucleotide sequence with at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1 and that encodes a polypeptide that is pesticidal for at least one pest belonging to the order Coleoptera. Thus, in addition to the steps required to practice independent claim 1, independent claim 9 requires the transformation of a plant. Plant transformation is routine in the art; thus, the amount of experimentation required to practice claim 9 is not undue. Similarly, in addition to the steps required to practice independent claim 1, the method of independent claim 17 requires that a nucleotide construct be created in which the nucleotide sequence is operably linked to a promoter; that the construct be introduced into a plant or cell thereof; and that an insect pest feeding on said plant or cell is impacted. The performance and/or evaluation required by each of these additional steps is within the skill of those in the art and would not be considered undue experimentation by those in the art. Likewise, the remaining sequence identity claims, which are all dependent on or incorporate the

limitations of independent claim 1, 9, or 17, contain additional requirements which are equally within the skill of those in the art.

Applicants note that it is now customary in the art to make and assay a number of sequences for a desired function in order to achieve the best results. For example, common techniques involve what is often referred to as "shuffling," as described for example in U.S. Patent No. 5,837,458, issued November 17, 1998 with inventors Minshull and Stemmer and entitled, "Methods and Compositions for Metabolic and Cellular Engineering." The Office Action (10/21/2004, page 9, first full paragraph) dismisses the teachings of this patent, stating that it "does not teach how to produce nucleic acids with specific identity to a known sequence." Applicants are confused by this statement, as the reference was cited to support Applicants' assertion that it is now customary in the art to make and assay a number of sequences for a desired function in order to achieve the best results, which is most definitely taught by the cited reference. In short, as illustrated by work described in U.S. Pat. No. 5,837,458, one of skill in the art would be able to produce novel sequences and evaluate whether they met the sequence identity limitations of the claims, as taught in the specification. One of skill in the art would then be able to identify whether those sequences retained pesticidal activity as taught in the specification. With "shuffling" techniques, it is common to mutagenize individual sequences or a set of sequences which are then assayed for a desired activity. Such techniques may even make use of a library of sequences which is recursively mutagenized, screened for function using a functional assay, and re-mutagenized in order to find a sequence exhibiting optimal function. Examples of the use of such techniques include: Minshull and Stemmer (1999) *Current Opinion in Chemical Biology* 3:284-290, entitled "Protein Evolution by Molecular Breeding"; and Christians *et al.* (1999) *Nature Biotechnology* 17: 259-264, entitled "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling." The Office Action (10/21/2004, page 9, second paragraph) stated that these references could not be considered because they were not sent; accordingly, these references are provided herewith as Appendix D and Appendix E. Reconsideration of these arguments is respectfully requested in view of the teachings of these references.

Such experiments are designed and are intended to encompass the generation and testing of a very large number of variant sequences for a desired function. As indicated by these and other publications in the art, this level of experimentation is now considered routine in the art and thus would not be considered "undue experimentation" under *In re Wands,* 858 F.2d 731, 8 USPQ2d 1400 (Fed Cir 1988) and *In re Jackson*, 217 USPQ 804, 807 (Bd. Pat. App. & Int. 1982) (holding that a considerable amount of experimentation is permitted to practice the invention and is not undue if it is merely routine in the art or if the specification provides a reasonable amount of guidance and direction to perform such experimentation).

The enablement of the claimed invention is not negated by the possibility of inoperative embodiments

The Office Action further states (10/21/2004, page 4, fourth paragraph) that

> The specification on pg 28, lines 5-11, suggests making these nucleic acids by making conservative substitutions in the encoded protein. However, making "conservative" substitutions (*e.g.*, substituting one polar amino acid for another, or one acidic one for another) does not produce predictable results. Lazar et al (1988, Mol. Cell. Biol. 8: 1247-1252) showed that the "conservative" substitution of glutamic acid for aspartic acid at position 47 reduced biological function of transforming growth factor alpha while "nonconservative" substitutions with alanine or asparagines had no effect (abstract). Similarly, Hill et al (1988, Biochem. Biophys. Res. Comm. 244: 573-577) teach that when three histidines that are maintained in ADP-glucose pyrophosphorylase across several species are substituted with the "nonconservative" amino acid glutamine, there is little effect on enzyme activity, while the substitution of one of those histidines with the "conservative" amino acid arginine drastically reduced enzyme activity (see Table 1). The nucleic acids encoding all these mutated proteins, however, would have much greater than 90% identity to the nucleic acids encoding the original protein.

It is true that some embodiments of the nucleotide sequence which meet the percent identity limitation of the claims may not encode a polypeptide that has the specified pesticidal activity. However, one of skill would readily be able to use the assays taught in the specification to determine which nucleotide sequences that met the sequence identity limitations of the claims also encoded polypeptides having the specified pesticidal activity. Applicants note that the

presence of inoperative embodiments within the scope of the claims does not render the claims invalid. *Atlas Powder Co. v. E.I. duPont de Nemours & Co.*, 750 F.2d 1569, 224 USPQ 409 (Fed. Cir. 1984). Nor would the amount of experimentation required to test a particular polypeptide for pesticidal activity be considered undue by one of skill in the art, as evidenced by the assay results presented in the specification, for example, in working Examples 4 (pp. 65-66), 6 (p. 67), and 7 (p. 69), and as evidenced by the Rule 132 declaration of Dr. André Abad submitted herewith as Appendix C. Indeed, the references cited in the Office Action—Lazar *et al.* (1988) *Mol. Cell. Biol.* 8: 1247-52 and Hill *et al.* (1998) *Biochem. Biophys. Res. Comm.* 244: 573-577—illustrate that one of skill would readily be able to determine whether a particular sequence change affected the function of a protein. Accordingly, one of skill in the art would be able to determine the functionality of polypeptides encompassed by the claimed invention without undue experimentation.

The Federal Circuit has repeatedly stated that enablement is not precluded by the necessity for some experimentation, so long as the experimentation needed to practice the invention is not undue. *In re Wands*, 858 F.2d 731 (Fed Cir 1988). Furthermore, a considerable amount of experimentation is permissible, if it is merely routine, or if the specification provides a reasonable amount of guidance in which the experimentation should proceed. *Id.* Applicants stress that when evaluating the quantity of experimentation required, the court looks to the amount of experimentation required to practice a single embodiment of the invention rather than the amount required to practice every embodiment of the invention as the Examiner implies. For example, in *Wands*, the claims at issue were drawn to immunoassay methods using any monoclonal antibody having a binding affinity for HbsAg of at least 10^{-9} M. The PTO had taken the position that the claim was not enabled because it would take undue experimentation to make the monoclonal antibodies required for the assay. The Federal Circuit reversed and held that the claims were enabled, as the amount of experimentation required to isolate monoclonal antibodies and screen for those having the correct affinity was not undue. *See Id.* Clearly, the Federal Circuit did not contemplate that every antibody useful in the methods of the claim must be identified. Rather, the court considered the amount of experimentation required to identify one or a few monoclonal antibodies having the required affinity. *See also, Johns Hopkins University*

v. Cellpro, 931 F. Supp. 303, 324 (D. Del. 1996), *aff'd in part, vacated in part, and remanded*, 47 USPQ2d 1705 (Fed. Cir. 1998) (stating that "[t]he specification need only enable one mode of making the claimed invention.").

Thus, for the reasons discussed above, Applicants respectfully submit that the sequence identity claims meet the enablement requirement of 35 U.S.C. §112, first paragraph. Based on the knowledge in the art and the guidance provided in the specification, the skilled artisan could choose among possible sequence modifications to produce polypeptides within the sequence identity parameters set forth in the claims and then test these sequence variants to determine if they retained pesticidal activity. The amount of experimentation needed to perform such an evaluation would not be considered by those of skill in the art to be undue; therefore, the amount of guidance presented in the specification is sufficient to enable the claims. Accordingly, Applicants respectfully submit that the Examiner's rejection of the sequence identity claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 under 35 U.S.C. §112, first paragraph, for lack of enablement should be withdrawn and should not be applied to new claim 65.

The Office Action (10/21/2004, page 10, #5) has rejected claims 1-3, 9-12, 17-19, 38, 43, 46, 49, 52, and 55-64 under 35 U.S.C. §112, first paragraph, as failing to meet the possession requirement "for the reasons of record as set forth in the Office action mailed 3 December 2003...." Applicants respectfully disagree with this rejection and the articulated bases for this rejection, and further note that this issue was on appeal in this case before prosecution was reopened and was discussed thoroughly in the Appeal Brief. In maintaining this rejection, the Examiner disregards not only Applicants' arguments but also the case law cited in those arguments. Applicants respectfully submit that the Examiner is applying an extraordinarily high standard of written description to the present claims, a standard that is not properly based on case law or on the statute.

Applicants note that the final Office Action (12/03/03, page 4, #4, 3d paragraph) stated that "nucleic acids that have 90% identity to SEQ ID NO:1 ***are predictable***, nucleic acids that have 90% identity to SEQ ID NO:1 AND that encode pesticidal proteins are not" (emphasis

added). Thus, the written description rejection is on the grounds that there is inadequate description of sequences that both meet the sequence identity requirement of the claims and also meet the functional requirement (*i.e.*, that the encoded polypeptide has pesticidal activity).

The claims meet the written description requirement as articulated by the Federal Circuit

Applicants respectfully submit that the present claims and specification meet the written description requirement of 35 U.S.C. §112, first paragraph, as clarified by *University of California v. Eli Lilly and Co.*, 119 F.3d 1559, 1569 (Fed. Cir. 1997) and *Amgen Inc. v. Chugai Pharmaceutical Co.*, 927 F.2d 1200 (Fed. Cir. 1991); *cert. denied* 112 S.Ct. 169 (1991). Applicants have provided exemplary sequences of the invention as set forth in SEQ ID NO:1. Indeed, the Office Action (10/21/2004, page 14, #9) indicates that claims limited to nucleotide sequences encoding the amino acid sequence set forth in SEQ ID NO:2 or having the nucleotide sequence set forth in SEQ ID NO:1 (*i.e.*, claims 39, 40, 44, 45, 50, and 51) would be allowable if rewritten in independent form. The claimed nucleic acids in the remaining claims are defined in relation to the exemplary disclosed nucleotide sequence of SEQ ID NO:1; that is, the claimed nucleic acids comprise nucleotide sequences that share a specified percentage of sequence identity with SEQ ID NO:1. Applicants have thus provided a structural definition of the sequences of the invention. Applicants have also provided assays by which those of skill in the art can readily assess whether a nucleic acid molecule meeting the nucleotide sequence element of the claims also meets the functional limitation element of the claims. This is what *Eli Lilly* requires, and Applicants have also conceived the sequences of the invention as articulated in *Amgen*; that is, Applicants are able "to envision the detailed constitution of a gene so as to distinguish it from other materials, as well as a method for obtaining it." *Amgen*, 927 F.2d at 1206.

Applicants further note that the Federal Circuit has explicitly stated that:

> *Eli Lilly* did not hold that all functional descriptions of genetic material necessarily fail as a matter of law to meet the written description requirement; rather, the **requirement may be satisfied if in the knowledge of the art the disclosed function is sufficiently correlated to a particular, known structure**.

(emphasis added; *Amgen, Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1332 (Fed. Cir. 2003). *See also, Moba, B.V. v. Diamond Automation, Inc.*, 325 F.3d 1306, 1320 (noting that "[i]n more recent cases, however, this court has distinguished *Lilly*" and further noting that in *Enzo Biochem, Inc. v. Gen-Probe, Inc.*, 323 F.3d 956 (Fed. Cir. 2002), "neither the specification nor the deposited biological material recited the precise 'structure, formula, chemical name, or physical properties' required by *Lilly*.")

B. thuringiensis δ-endotoxins are well-known in the art, and further support is provided in the specification with working examples

As discussed extensively in the specification (*e.g.*, on pp. 3, 7, 11-12, 15, 24-25), the disclosed exemplary sequence of SEQ ID NO:2 is a *Bacillus thuringiensis* Cry-8-like δ-endotoxin. The *B. thuringiensis* δ-endotoxins are an extremely well-characterized group of proteins. As discussed in the specification at pp. 24-25:

> The inventors of the present invention used the **solved structure of the *Cry3A* gene** (Li *et al.* (1991) *Nature* 353:815-821) to produce a homology model of the Cry8 δ-endotoxin disclosed and claimed herein as SEQ ID NO:2 to gain insight into the relationship between structure and function of the endotoxin, and to design the recombinantly engineered proteins disclosed and claimed herein. A combined consideration of the **published structural analyses** of *B. thuringiensis* endotoxins and the reported function associated with particular structures, motifs, and the like indicates that specific regions of the endotoxin are correlated with particular functions and discrete steps of the mode of action of the protein. For example, **δ-endotoxins isolated from *B. thuringiensis* are generally described as comprising three domains, a seven-helix bundle that is involved in pore formation, a three-sheet domain that has been implicated in receptor binding, and a beta-sandwich motif** (Li *et al.* (1991) *Nature,* 305: 815-821).

As discussed in more detail in the specification on page 25, the inventors made use of this knowledge to design specific mutations in the Cry8-like proteins to enhance their pesticidal activity. This strategy was successful in creating altered endotoxins with increased toxicity, as demonstrated by the data presented in working Example 6. Thus, having identified the novel sequences disclosed in the specification, the inventors were able, in view of the extensive knowledge in the art about *B. thuringiensis* δ-endotoxins, to modify these exemplary sequences

to provide an endotoxin with enhanced pesticidal activity. In this manner, the data provided in the specification (*e.g.*, Example 6) demonstrate that one of skill in the art would be able to determine which amino acids were more likely or less likely to disrupt the function of the protein when modified, and would be able to assay the protein produced to determine whether the modified protein retained pesticidal activity.

Additional knowledge in the art supports the description of the sequences of the claims. In addition to the Li reference's teaching of the tertiary structure of an exemplary Cry endotoxin, those of skill in the art are aware of conserved regions of the Cry endotoxins. Provided herewith as Appendix B is an alignment of exemplary SEQ ID NO:2 with the Pfam consensus domains for endotoxins, which are referred to as "Endotoxin N," "Endotoxin M," and "Endotoxin C" (Pfam accession numbers PF03945, PF00555, and PF03944, respectively, descriptions of which are also attached as Appendix B). The Pfam database provides a curated collection of well-characterized protein family domains with high quality alignments. It is well known in the art that regions of sequence homology with known functional domains may be used to determine protein function and to identify what regions of a protein are particularly conserved (and therefore less likely to tolerate mutations) as well as what regions of a protein are less conserved (and therefore more likely to tolerate mutations). Therefore, Applicants respectfully submit that the novel exemplary sequences disclosed in the specification, combined with the knowledge of one familiar with the art, provide adequate guidance as to which amino acids of SEQ ID NO:2 may be modified while allowing the protein to retain pesticidal activity. Accordingly, Applicants respectfully submit that they have envisioned the detailed construction of the gene to distinguish it from other materials and so that one of skill in the art would recognize that Applicants were in possession of the claimed subject matter at the time the application was filed, thereby meeting the written description requirement.

Applicants note that the Office Action (10/21/2004, page 12, fourth paragraph and last paragraph) continues to dismiss the teachings in the specification of sequences that are exemplary variants and fragments of SEQ ID NO:1 and SEQ ID NO:2. Applicants emphasize that the specification teaches a number of nucleic acids that share relatively low percent sequence identity with SEQ ID NO:1 but encode proteins having pesticidal activity. In Example

4 (specification pp. 65-66), both the full-length endotoxin encoded by SEQ ID NO:1 and a truncated protein encoded by SEQ ID NO:15 were assayed for pesticidal activity against southern corn rootworm. The nucleotide sequence of SEQ ID NO:15 is a truncation of SEQ ID NO:1 which shares about 55% sequence identity with SEQ ID NO:1. In Example 6 (specification pp. 67-69), several truncated proteins were assayed and shown to have pesticidal activity against Colorado potato beetle (see Table 1, p. 68). These truncated proteins included those encoded by SEQ ID NO:15 and SEQ ID NO:19, which share about 55% and 51% identity, respectively, to the exemplary nucleic acid sequence set forth in SEQ ID NO:1 (alignments performed using BLAST with default parameters). Another mutant assayed for pesticidal activity in Example 6 was NGSR1218-1 (encoded by SEQ ID NO:11). The NGSR1218-1 mutant includes the amino acid sequence "NGSR" inserted between amino acids 164 and 165 of the truncated endotoxin of SEQ ID NO:16. The nucleotide sequence encoding this mutant (SEQ ID NO:11) shares about 56% sequence identity with the exemplary nucleotide sequence of SEQ ID NO:1, yet both encoded proteins have pesticidal activity. The specification also teaches an exemplary maize-optimized sequence (SEQ ID NO:9), which encodes the same pesticidal polypeptide as SEQ ID NO:15 but shares less than 69% sequence identity with it. Thus, the present specification provides multiple working examples illustrating the production of sequences that encode pesticidal proteins and share a relatively low percentage of sequence identity with SEQ ID NO:1. Multiple working examples are presented, illustrating that Applicants were in possession of the claimed invention at the time of filing.

The Office Action (10/21/2004, page 12, fourth paragraph and last paragraph) continues to dismiss these teachings of the specification, stating, for example, that "the local match similarity between SEQ ID NO:15 and the first half of SEQ ID NO:1 is 100%." Applicants respectfully disagree with this assessment of the significance of the examples in the specification and note that these truncations, along with the knowledge in the art discussed above, help to clarify the importance of different portions of the exemplary proteins so that one of skill would recognize which portions of the exemplary proteins are more likely to be necessary for pesticidal activity and which portions of the exemplary proteins are less likely to be important for such activity. In light of the above statements, Applicants respectfully submit that the present claims

and specification satisfy the statutory written description requirement. Accordingly, Applicants respectfully request that the rejection of the sequence identity claims under 35 U.S.C. §112, first paragraph, be withdrawn and not be applied to new claim 65.

The Rejection of Claims Under 35 U.S.C. §103(a) Should Be Withdrawn

The Office Action (10/21/2004, page 13, #7) has rejected claims 1-3, 9-12, 17-19, 46, 52, 57, and 60-62 under 35 U.S.C. §103(a) as being unpatentable over Michaels *et al.* (U.S. Pat No. 5,554,534). Applicants respectfully traverse this rejection.

Independent claim 1 is drawn to an isolated nucleic acid comprising a nucleotide sequence having at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1, wherein said nucleotide sequence encodes a polypeptide which is pesticidal for at least one pest belonging to the order *Coleoptera*.

The Office Action (page 13) characterizes the Michaels reference as disclosing "the sequence of a scarab-specific Cry protein, their SEQ ID NO:4, which is pesticidal for scarabs belonging to the genus Cotinus (column 14, lines 26-44) and claim all nucleic acids encoding that protein (claim 1). Scarabs belong to the order Coleoptera; thus, [Michaels'] SEQ ID NO:4 is pesticidal for at least one pest belonging to the order Coleoptera" as presently claimed.

As Applicants understand it, the rationale for this obviousness rejection is as follows. The Office Action notes (page 5) that the instant SEQ ID NO:1 is 3621 nucleotides in length and then states:

> [B]ecause nucleic acids that have 90% identity to SEQ ID NO:1 would have up to 362 nucleotide substitutions, they could encode proteins with up to 362 amino acid substitutions; these proteins would have 70% identity to the 1206 amino acid long SEQ ID NO:2. * * *
>
> [As described in the specification,] SEQ ID NO:1 has homology to GenBank U04365, which is identical to SEQ ID NO:3 of Michaels et al.... This nucleotide sequence has 85.1% identity to SEQ ID NO:1 [and so does not fall within the genus claimed in claim 1]; however, it encodes a protein with 79.8% identity to the instant SEQ ID NO:2....

The Office Action concludes (page 14):

> Michaels et al [*sic*] do not disclose an individual nucleic acid with 90% identity to the instant SEQ ID NO:1. At the time the invention was made, it would have been obvious to one of ordinary skill in the art to [*sic*] that nucleic acids with 90% identity to SEQ ID NO:1 are encompassed within the full scope of nucleic acids encoding the scarab-specific Cry protein taught by Michaels et al.

Thus, the Office Action rejects claims drawn to nucleotide sequences that have at least 90% sequence identity to SEQ ID NO:1. As set forth in more detail below, Applicants respectfully disagree with this rejection and with the rationale asserted for the rejection. However, as an initial matter, Applicants believe that the Office Action is somewhat inconsistent in its application of this rejection to claims with different sequence identity limitations. Particularly, the Office Action concedes (page 14) that:

> Claims 38-40, 43-45, 49-51, and 55-59 are free of the prior art, given the failure of the prior art to teach or suggest an isolated nucleic acid of SEQ ID NO:1 or encoding SEQ ID NO:2, or as [*sic*] isolated nucleic acid with 93% identity to SEQ ID NO:1, wherein the nucleic acid encodes a protein pesticidal for at least one pest belonging to the order Coleoptera.

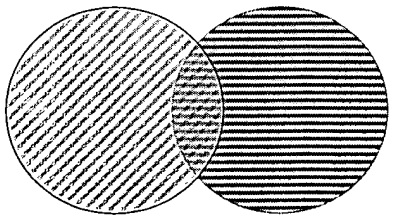
It would seem from this statement that claims 63 and 64 (with limitations to 93% and 94% sequence identity, respectively) should also have been included in the list of claims free of the prior art. Moreover, if Applicants understand the basis for the rejection, it would seem that the rule applied in making the rejection would have resulted in the determination that claims with limitations to at least 94% sequence identity and above avoid the cited art, rather than claims with limitations to at least 93% sequence identity and above. The rule applied in making the rejection appears to be as follows: Let x be the difference between 100% and the percentage of nucleotide sequence identity between two sequences (for claim 1, x = (100% - 90%) = 10%). Then y is the difference between 100% and the percentage of amino acid sequence identity of the least similar possible proteins encoded by the nucleotide sequences in question, and $y = 3x$.

To illustrate this calculation for the cited prior art in the reverse direction (*i.e.*, to determine what nucleotide sequences might be encompassed by the Michaels claim): proteins that share 80% sequence identity can be encoded by nucleotide sequences that share x% sequence identity, and to find x, the calculation is: y = (100% - 80%) = 20%; $y/3 = x$, so x =

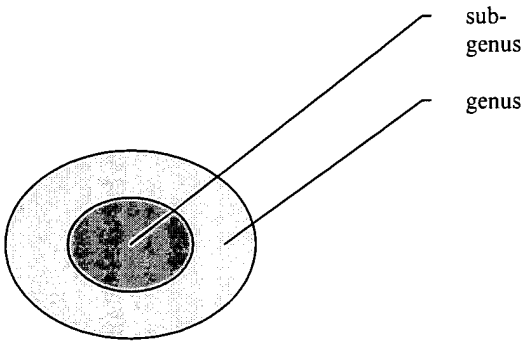
6.67% and the nucleotide sequences that encode these least similar possible proteins could share as little as (100% - 6.67% =) 93.33% sequence identity. So here, following the rule that seems to have been applied in the Office Action (and with which Applicants disagree), the Office Action on that basis should have limited the claims free of the art to claims specifying 94% sequence identity and above, and should have included claim 64 in the list of claims deemed to be free of the prior art.

Returning to the rejection of claims and the rationale asserted for this rejection, Applicants respectfully disagree with both. However, in researching this issue, it does not appear that this precise question is discussed in the MPEP, nor does it seem to have been considered by a court or the Board of Appeals. In trying to understand whether the rejection might be upheld, Applicants turned to similar fact patterns in the MPEP and in the case law. It is clear that a disclosure in the prior art of a species falling within a genus anticipates that genus (see MPEP §2131.02). However, here, as acknowledged in the Office Action, no such species were described, and so the rejection was made on the basis of obviousness.

To appreciate the differences between the present situation and those addressed in the MPEP and in the case law, consider a Venn diagram (not necessarily to scale) of the nucleotide sequences encompassed by the cited art genus and the nucleotide sequences encompassed by the presently claimed genus. These genera overlap, and the diagram would look something like this:



The MPEP does not appear to directly address the situation of overlapping genera. The most similar situation discussed in the MPEP of which Applicants are aware is the obviousness of species when the prior art teaches a genus (MPEP §2144.08). This section discusses situations where a claimed species or subgenus is encompassed by a genus taught in the prior art; a Venn diagram of this situation would look something like this:



Clearly, these two situations differ. However, the closest situation discussed in the MPEP appears to be the discussion of genus/species analysis in MPEP §2144.08. Most of the statements in this section are general statements about obviousness, but they are reiterated in this section in the context of genus/species analysis, as follows. Determinations of patentability under 35 U.S.C. §103 should be made on the facts of each case "in the totality of the circumstances." The test is "whether the claimed species or subgenus would have been obvious to one of ordinary skill in the pertinent art at the time the invention was made." In order to establish a *prima facie* case of obviousness "in a genus-species chemical composition situation, ... ***it is essential that Office personnel find some motivation or suggestion to make the claimed invention in light of the prior art teachings***" (citing *In re Brouwer*, 77 F.3d 422, 425 (Fed. Cir. 1996)). Where a motivation or suggestion to make the claimed invention is asserted, "there should be a reasonable likelihood that the claimed invention would have the properties disclosed by the prior art teachings. The prior art disclosure may be express, implicit, or inherent" (citing

In re Vaeck, 947 F.2d 488, 493). "***The fact that a claimed species or subgenus is encompassed by a prior art genus is not sufficient by itself to establish a prima facie case of obviousness***" (emphasis added; citing *In re Baird*, 16 F.3d 380, 382 (Fed. Cir. 1994)).

These statements in the MPEP contradict the conclusion reached in the Office Action, because while the cited art sequence does share a high degree of sequence identity with the present SEQ ID NO:1, the cited art does not teach or suggest all the limitations of any of the presently pending claims and therefore the claims are not obvious in view of the reference.

Turning to case law, Applicants have searched for but were unable to find any case in which this issue had been squarely addressed. The two most relevant cases are the well-known Federal Circuit cases *In re Bell* and *In re Deuel*, which are considered in detail below.

In *In re Bell*, the Federal Circuit reversed the rejection of claims for obviousness. The claims at issue in *In re Bell* (991 F.2d 781 (Fed. Cir. 1993)) differ from the present claims because Bell's claims were narrowly drawn to particular sequences. Particularly, Bell's claims were drawn to nucleic acid molecules (DNA and RNA) encoding human insulin-like growth factors I and II (IGF-I and IGF-II). Representative claim 25 was drawn to "[a] composition comprising nucleic acid molecules containing a human sequence encoding [human IGF]...wherein said hIGF sequence is selected from the group consisting of: (a) [a first specific sequence]; (b) [a second specific sequence]; (c) nucleic acid sequences complementary to (a) or (b); and (d) fragments of (a), (b), or (c) that are at least 18 bases in length and which will selectively hybridize to human genomic DNA encoding hIGF." The cited prior art included two publications that disclosed the amino acid sequences for IGF-I and IGF-II as well as a patent (U.S. Pat. No. 4,394,443 to Weissman *et al.*) that described a general method for isolating a gene where at least a partial amino acid sequence of the encoded protein was known.

In reversing the rejection of the claims for obviousness, the Federal Circuit concluded that ***where the prior art suggested a vast number of possible nucleic acid sequences, the claimed individual sequences would not have been obvious***. The court also noted and disagreed with the PTO's argument that a gene is rendered obvious once the amino acid sequence of its translated protein is known. The court emphasized that the PTO's analysis of obviousness improperly focused on the similarity of Bell's methods to the prior art methods because Bell did

not claim methods but rather claimed compositions. The court also noted, in *dicta*, that Bell did not claim all of the nucleic acids that might potentially code for IGF or all nucleic acids encoding a protein having the same biological activity.

In *In re Deuel* (51 F.3d 1552 (Fed. Cir. 1995)), the PTO's rejection of claims for obviousness was reversed by the Federal Circuit. *In re Deuel* involved claims to particular nucleotide sequences as well as genus claims that were similar to the present claims, but the court did not consider the genus claims separately because the rejection of Deuel's claims grouped the genus claims together with the claims to individual sequences. Deuel had isolated and purified Heparin-Binding Growth Factor (HBGF) from bovine uterus tissue. In addition, he had found that HBGF stimulated cell division, had identified the first 25 amino acids of the protein, had cloned and sequenced the cDNA, and had predicted the complete amino acid sequence of bovine uterine HBGF. Deuel had also isolated the human HBGF cDNA, determined its nucleotide sequence, and determined the predicted amino acid sequence of the human HBGF protein. Deuel's independent claims 5 and 7 were drawn to cDNAs of human and bovine HBGF, respectively, wherein each cDNA had a particular nucleotide sequence. Independent claims 4 and 6 were drawn to cDNAs *encoding* human HBGF protein and bovine HBGF protein, respectively, wherein each protein had a particular amino acid sequence.

The "Bohlen" reference (EP App. No. 0326075) cited against Deuel's claims taught a group of brain-specific proteins which were named "Heparin-Binding Brain Mitogens" ("HBBMs") that were useful in promoting the growth and repair of neural tissue; Bohlen had determined the first 19 amino acids of these proteins, which were identical for both human and bovine HBBMs and which matched the first 19 amino acids of Deuel's HBGF proteins. No cDNA or other nucleotide sequence was disclosed. The Examiner asserted that Deuel's HBGF proteins would have been obvious in view of the N-terminal sequence of the HBBMs and general cloning methods known in the art. The Bohlen reference suggested that HBBMs may be homologous between species, but did not indicate that there could be homology between tissue types.

The court, analyzing the particular species disclosed in claims 5 and 7, concluded that "the ***precise*** cDNA molecules of Deuel's claims 5 and 7 ***would not have been obvious*** over the

Bohlen reference ***because Bohlen teaches protein***, not the claimed or closely related cDNA molecules." The court discussed that because of the redundancy of the genetic code, the amino acid sequence of the protein did not lead one of skill in the art to contemplate or focus on the specific cDNAs of claims 5 and 7. As in *In re Bell*, the court emphasized that the PTO's focus on the methods used to identify the compounds was improper because the claims were drawn to compositions, not methods: "[t]he fact that one can conceive a general process in advance for preparing an *undefined* compound does not mean that a claimed specific compound was *precisely* envisioned and therefore obvious" (emphasis in original). The court held that "[w]e today reaffirm the principle, stated in *Bell*, that ***the existence of a general method of isolating cDNA or DNA molecules is essentially irrelevant to the question whether the specific molecules themselves would have been obvious, in the absence of other prior art that suggests the claimed DNAs***." The court concluded that "[u]ntil the claimed molecules were actually isolated and purified, it would have been highly unlikely for one of ordinary skill in the art to contemplate what was ultimately obtained. What cannot be contemplated or conceived cannot be obvious."

Deuel and *Bell* were decided some time ago. However, the approach in *Deuel* and *Bell* to DNA and protein sequences was reemphasized recently in *In re Wallach* (378 F.3d 1330 (Fed. Cir. 2004)). Applicants had isolated two proteins, Tumor Necrosis Factor Binding Proteins I and II (TBP-I and TBP-II) and had determined the proteins' molecular weight, their elution characteristics from reverse-phase HPLC, and partial amino acid sequences (ten amino acids in length). Representative claim 11 was drawn to an isolated DNA comprising a nucleotide sequence that encoded human TBP-II, wherein: the protein's amino acid sequence included the identified partial sequence of ten amino acids; the protein could inhibit the cytotoxic effect of TNF; the protein eluted in a particular reversed-phase HPLC fraction, and the protein had a molecular weight of about 30kDa.

The Federal Circuit affirmed the Board's determination that the claims had been properly rejected as failing to meet the written description requirement. In making this determination, the court elaborated on the standard applied to nucleotide and amino acid sequences. The court noted that "the state of the art has developed such that the *complete* amino acid sequence of a

protein may put one in possession of the genus of DNA sequences encoding it...," (emphasis added) and it is "a routine matter to convert back and forth between an amino acid sequence and the sequences of the nucleic acid molecules that can encode it."

The *Wallach* opinion reiterated the distinction between individual sequences and a genus of sequences as well as the distinction between support and description. The opinion discussed that the support necessary for a genus (*i.e.*, the description of a representative number of species in a genus) does not require individual support for each species encompassed by that genus: "for example,...a disclosure of an amino acid sequence would provide sufficient information such that one would accept that an applicant was ***in possession of the full genus of nucleic acids encoding a given amino acid sequence, but not necessarily any particular species***" (citing MPEP §2163.II.A.3.a.ii (8th Ed., rev. 2 (2001)). Therefore, the court concluded, Wallach might have been in possession of the entire genus of DNA sequences that could encode the disclosed partial protein sequence, "***even if individual species within that genus might not have been described or rendered obvious***" (citing *In re Deuel*, 51 F.3d 1442 (Fed. Cir. 1995)). The court concluded that Wallach's application had not met the written description requirement because he had only provided a *partial* protein sequence and had not provided evidence of any correlation between this property and the structure of the DNA encoding the protein.

Returning to the present issue of overlapping genera, it is unfortunate that none of these cases—*In re Bell*, *In re Deuel*, and *In re Wallach*—squarely address this issue. However, these cases do provide some guidance regarding the analysis of genera and species. Particularly, the case law clearly distinguishes between the *possession* of a genus of sequences and the description of that genus. Here, the Michaels reference may show *possession* of the genus of nucleotide sequences encoding Michaels' claimed protein, but it did not describe any individual species or any subgenus that would anticipate the presently claimed genus. The key question is, how can a claimed genus ***as a whole*** be considered obvious if only a small portion of the genus was disclosed?

Phrased a little differently, the question here is whether the genus of sequences disclosed in the Michaels reference renders obvious the presently claimed genus solely because those

genera "overlap" by virtue of containing some of the same species. Applicants emphasize that here, none of the species encompassed by either genus were individually described in the prior art. Further, the portion of the claimed genus within the overlapping region was not described in the art as a discrete subgenus, but was only present in the cited reference as an indiscrete portion of a larger genus. Thus, there is no anticipation of the claimed genus by earlier disclosure either of a species or of a subgenus, and the Office Action (10/21/2004, page 14, #8) has acknowledged that at least part of the genus is free of the prior art.

Further, Applicants emphasize that the claimed genus is defined by reference to a novel and nonobvious nucleotide sequence. It is well settled that all of the claim limitations must be considered in evaluating obviousness. As summarized in MPEP §2142:

> To establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest ***all*** the claim limitations.

(emphasis added). Here, the cited art does not teach or suggest all of the claim limitations because it does not teach the sequence of SEQ ID NO:1. Therefore, the claimed genus can not be obvious in view of Michaels, and it must also be true that a genus that is defined by reference to the structure of a previously undisclosed sequence is not obvious. Applicants are mindful that the present situation is one where the cited art genus and the claimed genus overlap only in a relatively small area; it is conceivable that the conclusion would differ where the genera overlapped in a relatively large area.

Finally, Applicants would like to address the assumption in the Office Action that the presently claimed nucleotide sequence of SEQ ID NO:1 could be changed by 10% so as to encode the protein disclosed in the Michaels reference (see, *e.g.*, statement on page 13, #7, first paragraph, that nucleic acids with 90% identity to SEQ ID NO:1 would encode proteins with up to 70% identity to the instant SEQ ID NO:2 and the statement on page 14, first paragraph, that the "protein taught by Michaels et al has 79.8% sequence identity to the instant SEQ ID NO:2..."). On further analysis as described below, Applicants believe that their exemplary SEQ

ID NO:1 would have to be changed by more than 21% to encode the amino acid sequence of Michaels' SEQ ID NO:3. Thus, while the genera of nucleotide sequences encompassed by these two genera seem to overlap conceptually, there is no nucleotide sequence within the claimed genus that could actually encode the Michaels protein.

In order to determine whether a nucleotide sequence within the claimed genus could encode the Michaels protein, Applicants asked two bioinformaticians to analyze the question. The bioinformaticians developed a set of scripts that use the GAP alignment of SEQ ID NO: 2 compared to SEQ ID NO: 3 (see, *e.g.*, specification pp. 34-35) to drive the calculation of the minimal required nucleotide changes and insertions and deletions to allow inter-conversion between the sequences. (See Rule 132 declarations of Dr. Carl Simmons and Dr. Thomas Z. McNeill, submitted herewith as Appendix F and Appendix G.) The appropriate metric to calculate the global distance between two sequence strings is called a Levenshtein distance (string edit distance). Theoretically one could generate all possible nucleotide sequences of the reference sequence and determine the sequences(s) that have the smallest Levenshtein distance between it and the claimed nucleotide sequence, but in practice this would create an enormous amount of possible sequences and a very complicated computational analysis.

To simplify the problem, the bioinformaticians used the GAP alignment of SEQ ID NO: 2 to that of the Michaels protein (Genbank U04365) to calculate the smallest Levenshstein distance (least number of nucleotides to change) for each amino acid codon position (codon by codon), and then summed these differences across the protein open reading frame (ORF) and expressed the answer as a percent of the nucleotide ORF length. For this analysis, a scoring system was developed using the Levenshtein distance per each codon. Identical amino acids could be coded with the identical codons so they were automatically assigned a minimal distance score of 0. For aligned but mismatched amino acids, the codon of SEQ ID NO:1 that codes for the aligned amino acid of SEQ ID NO: 2 was compared against all possible codons for the aligned amino acid of Michaels protein (Genbank U04365), and the codon for that amino acid requiring the least number of nucleotide changes was determined. These values could range from 1 to 3 nucleotide changes, with the average nearest to 2 Gaps and insertions in the protein alignment of sequences 2 and 3 were automatically scored the largest possible distance or 3

nucleotide changes. These scores were then summed and divided by the number of nucleotides for SEQ ID NO: 1, the necessary SEQ ID NO: 2 ORF length, which yielded a minimal percentage of nucleotides that would need to be changed (altered, inserted or deleted) in order to convert between the claimed and the reference sequences in nucleotide space. The answer was 21.6%, which is much higher than the 10% variation specified in the claims. Therefore, by these methods there are no nucleotide sequences having 90% sequence identity to the sequence set forth in SEQ ID NO:1 that encode the Michaels protein discussed in the Office Action.

For the reasons discussed above, Applicants respectfully submit that because the cited art does not teach or suggest all of the claim limitations of any of the claims, the claimed genus is not rendered obvious in view of that art and the rejection of claims under 35 U.S.C. §103 should be withdrawn and should not be applied to new claim 65.

CONCLUSION

In view of the above amendments and remarks, Applicants submit that the rejections of the claims under 35 U.S.C. §112, first paragraph, and 103(a) are overcome. Applicants respectfully submit that this application is now in condition for allowance. Early notice to this effect is solicited.

If in the opinion of the Examiner, a telephone conference would expedite the prosecution of the subject Application, the Examiner is invited to call the undersigned.

It is not believed that extensions of time or fees for net addition of claims are required, beyond those that may otherwise be provided for in documents accompanying this paper. However, in the event that additional extensions of time are necessary to allow consideration of this paper, such extensions are hereby petitioned under 37 CFR § 1.136(a), and any fee required

therefore (including fees for net addition of claims) is hereby authorized to be charged to Deposit Account No. 16-0605.

Respectfully submitted,

Leigh W. Thorne
Registration No. 47,992

Customer No. 29122
ALSTON & BIRD LLP
Bank of America Plaza
101 South Tryon Street, Suite 4000
Charlotte, NC 28280-4000
Tel Raleigh Office (919) 862-2200
Fax Raleigh Office (919) 862-2260

"Express Mail" mailing label number EV 387068832 US
Date of Deposit January 21, 2005

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to:
Mail Stop Amendment, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Lynda-Jo Pixley

RTA01/2172746

APPENDIX A

42. Porter, N. A. & Weekes, T. C. *Mon. Not. R. astrr. Soc.* **183**, 205-210 (1978).
43. Akerlof, C. W. *et al.* in *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 135-138 (Adelaide University, 1990).
44. Porter, N. A. & Weekes, T. C. *Astrophys. J.* **212**, 224-226 (1977).
45. Porter, N. A. & Weekes, T. C. *Mon. Not. R. Astr. Soc.* **183**, 205-210 (1978).
46. Porter, N. A. & Weekes, T. C. *Nature* **277**, 199 (1979).
47. Nolan, K., Porter, N. A., Fegan, D. J., Chantell, M. & Weekes, T. C. *Proc. 21st ICRC, Adelaide*, vol. 2 (ed. Protheroe, R. J.) 150 (Adelaide University, 1990).
48. Fegan, D. J., McBreen, B., O'Brien, D. & O'Sullivan, C. *Nature* **271**, 731-732 (1978).
49. Bhat, P. N. *et al. Nature* **284**, 433-434 (1984).
50. Porter, N. A. & Weekes, T. C. *Nature* **267**, 500-501 (1977).
51. O'Mongain, E. *Nature* **242**, 136-137 (1973).
52. Huguenin, G. R. & Moore, E. L. *Astrophys. J.* **187**, L57-L58 (1974).
53. Phinney, S. & Taylor, J. H. *Nature* **277**, 117-118 (1979).
54. Rees, M. J. *Nature* **266**, 333-334 (1977).
55. Jelley, J. V., Baird, G. A. & O'Mongain, E. *Nature* **267**, 499-500 (1977).

ACKNOWLEDGEMENTS. We thank C. J. Goebel for giving his time whenever we needed expert help. We thank N. Porter for discussions. This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, in part by the US Department of Energy, in part by the Xunta de Galicia (Spain) and in part by the Smithsonian Scholarly Studies Research Fund. J.H.M. is a NAS/NRC research associate.

ARTICLES

Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution

Jade Li*, Joe Carroll† & David J. Ellar†

* Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Biochemistry Department, Cambridge University, Tennis Court Road, Cambridge CB2 1QW, UK

The structure of the δ-endotoxin from *Bacillus thuringiensis* subsp. *tenebrionis* that is specifically toxic to Coleoptera insects (beetle toxin) has been determined at 2.5 Å resolution. It comprises three domains which are, from the N- to C-termini, a seven-helix bundle, a three-sheet domain, and a β sandwich. The core of the molecule encompassing all the domain interfaces is built from conserved sequence segments of the active δ-endotoxins. Therefore the structure represents the general fold of this family of insecticidal proteins. The bundle of long, hydrophobic and amphipathic helices is equipped for pore formation in the insect membrane, and regions of the three-sheet domain are probably responsible for receptor binding.

THE δ-endotoxins are a family of insecticidal proteins produced by *Bacillus thuringiensis* (B.t.) during sporulation, having relative molecular masses (M_r) 60,000-70,000 (60K-70K) in the active form and specific toxicities against insects in the orders of Lepidoptera, Diptera and Coleoptera[1,2]. These toxins have been formulated into commercial insecticides for three decades[3], and now insect-resistant plants are engineered by transformation with Lepidoptera-specific toxin genes[4-6]. In the bacterium δ-endotoxins are synthesized as protoxins of M_rs 70K-135K and crystallize as a parasporal inclusion ~1 μ in size, in which form they are ingested by the susceptible insect. The microcrystal dissolves in the alkaline pH of the midgut and the protoxin is cleaved by gut proteases to release the active toxin. δ-Endotoxins activated *in vitro* bind specifically and with high affinity ($k_D \approx$ 0.1-20 nM) to protein receptors on brush-border membrane vesicles derived from the gut epithelium of target insects[7-9] and create leakage channels of 10-20 Å diameter in the cell membrane[10]. *In vivo* such membrane lesions lead to swelling and lysis of the gut epithelium[11] and death of the insect ensues through starvation and septicaemia. Active δ-endotoxins of different specificities show five strongly conserved regions in their amino-acid sequences[1,12]. Exchanging sequence segments in the divergent regions between toxins of different specificities can produce active hybrids showing altered target specificity[13-15]. We have determined the atomic structure of a Coleoptera-specific δ-endotoxin (CryIIIA, beetle toxin) from *B.t.* subsp. *tenebrionis*[16-18] to elucidate the structural basis for target specificity and membrane perforation by this family of proteins.

Structure determination

Parasporal crystals of the beetle toxin contain the full-length 644-residue protoxin[17] as the minor component, and a product of bacterial processing with 57 residues removed from the N-terminus as the major component[19]. The latter (M_r 67K) is similar in sequence to the active form of other δ-endotoxins. After solubilization, papain cleavage converts the mixture to the 67K toxin (see legend to Table 1). This was recrystallized in the original crystal form of the parasporal crystals, space group $C222_1$ and cell dimensions 117.1 by 134.2 by 104.5 Å, containing one molecule per asymmetric unit and 55% solvent by volume[18].

Initial evaluation of derivatives was carried out at 4.5 Å resolution with data collected on the FAST TV diffractometer[20] using CuKα radiation. Complete datasets (Table 1) were then collected to 2.5 Å resolution from native crystals using the imaging plate systems at the EMBL outstation at DESY and from the mercury and platinum derivatives on film at SRS Daresbury. The electron density map (Fig. 1) at 2.5 Å resolution calculated with phases from multiple isomorphous replacement (mean figure of merit, 0.63) was easily interpretable and was improved by solvent flattening[21,22]. A continuous polypeptide chain from residue 61 to residue 644 at the C terminus was traced unambiguously, and most side-chain atoms could be located in the map. The atomic model was built using the graphics program O (ref. 23) and had an initial *R*-factor of 37% for all data to 2.5 Å. After preliminary refinement using the program X-PLOR (ref. 24), the current model, containing 584 amino acid residues and 40 bound water molecules, has an *R*-factor of 19.9% and r.m.s. bond length deviation of 0.017 Å.

Description of the structure

Overview. The beetle toxin is a wedge-shaped molecule with a radius of gyration of 58 Å. As shown in Fig. 2*a*, it comprises three domains. Domain I, from the N terminus of the 67K toxin to residue 290, is a seven-helix bundle in which a central helix is completely surrounded by six outer helices tilted at about +20° to it (Fig. 3*b*,*c*). Domain II, from residues 291 to 500, contains three antiparallel β sheets packed around a hydrophobic core with a triangular cross-section (Fig. 4). Domain III, from residues 501 to 644 at the C terminus is a sandwich of two antiparallel β sheets (Fig. 5). Domains I and III make up the

TABLE 1 Data collection and phasing statistics

Data collection

Data	Method of collection	Number of crystals	Resolution (Å)	Number of measurements	Unique reflections (% completeness)	R_{merge}
Native	image plate	8	2.5	121,767	27,727 (100)	0.108
CH_3HgNO_3	film	7	2.5	103,623	27,767 (100)	0.095
$Hg(CH_3COO)_2$	film	5	2.5	60,224	25,919 (94.5)	0.103
cis-$Pt(NH_3)_2Cl_2$	film	7	2.5	86,629	25,924 (94.5)	0.107
K_2OsO_4	FAST	1	4.5	21,143	4,680 (100)	0.077
$HoCl_3$	FAST	1	4.5	20,013	4,701 (100)	0.069

Phasing statistics

Derivative	Anomalous data	Number of sites	R_{deriv}†	R_{Cullis}‡	Phasing power§ (resolution, Å)
CH_3HgNO_3	no	3	0.183	0.715	1.56 (2.5)
$Hg(CH_3COO)_2$	yes	6	0.247	0.609	2.28 (2.5)
cis-$Pt(NH_3)_2Cl_2$	no	5	0.185	0.682	1.54 (2.5)
K_2OsO_4	no	4	0.149	0.757	1.26 (5.5)
$HoCl_3$	no	3	0.095	0.741	1.35 (5.0)

Protein preparation: Solubilized parasporal crystals from *B.t.* subsp *tenebrionis* were incubated at 0.5 mg ml^{-1} protein with 0.125 units per ml Agarose-linked papain (Boehringer) in 3.3 M NaBr, 0.05 M sodium phosphate, pH 7.0, and 0.1 mg ml^{-1} phenylmethylsulphonylfluoride (PMSF) for 30 min 20 °C. Digestion was stopped by adding tosyl lysinechloromethylketone (TLCK) to 0.125 mg m^{-1} and Na_2CO_3 to one fifth volume and removing t enzyme-beads. The 67K beetle toxin was then purified by gel filtration on Sephadex G75 equilibrated with 0.1 M $NaHCO_3$, pH 10.5, 0.5 M NaBr. **Crystallizatic** Single crystals were obtained by microdialysis at a protein concentration of 2.5 mg ml^{-1} against 0.1 M $NaHCO_3$, pH 9.5, 1.2 M NaBr at 4 °C overnight, th against 0.1 M $NaHCO_3$, pH 9.2, 0.5 M NaBr at 16 °C; 3 mM NaN_3, 0.1 mM PMSF and 0.1 mg ml^{-1} TLCK were present in all buffers. Crystals were transferr by stages to 0.05 M 2-(*N*-morpholino)ethanesulphonic acid (MES), pH 6.5, for derivative preparation and mounted in 0.03% low-melting agarose in this buff during data collection. **Data collection:** Image plate and film data were processed using MOSFLM (Imperial College, London) and CCP4 programs (Daresbu UK). FAST (ref. 20) data were collected and processed with MADNES[45], and scaled in 3° batches. **Derivatives:** Crystals were soaked respectively in 0.25 m CH_3HgNO_3 for 3.5 h, in 1 mM $Hg(CH_3COO)_2$ for 14 h, in freshly prepared 1 mM *cis*-$Pt(NH_3)_2Cl_2$ for 21 h, in saturated K_2OsO_4 for 35 h, and in 2 mM $HoCl_3$ f 3 days. **Phase calculation:** Two heavy-atom sites in each derivative were located from difference Patterson functions, except in the case of $Hg(CH_3COO$ for which 3 sites were located, and the remaining sites were found by cross-phased difference Fouriers. Heavy-atom parameters were refined again centric data and phases calculated for all data using the program PHARE (G. Bricogne). The two low-resolution derivatives were refined against phase calculated from the high-resolution derivatives. Phasing with the three high-resolution derivatives gave an overall figure of merit of 0.61 (25–2.5 Å) and clearly interpretable map. Including the remaining derivatives slightly improved the connectivity of the map (overall figure of merit 0.63), and four cycles solvent flattening using a 50% solvent content and a 9 Å radius in mask calculation[21,22] improved the overall definition of densities. The starting mod was built using the program *O* (ref. 23) with the Bones option for main-chain tracing and the autobuild and manip options for side chains. Refinement b simulated annealing using the program X-PLOR (ref. 24) reduced the *R*-factor from 0.37 to 0.25 without individual *B*-factors, and to 0.23 with restraine individual *B*-factors. The model was adjusted in the loops 154–156, 429–436, and 483–488, and had 40 solvent molecules added, then refined by X-PLO again. The current model has an *R*-factor of 19.9%, with r.m.s. bond length deviation of 0.017 Å, r.m.s. bond angle deviation of 3.2°, and average atom *B*-factor of 18 Å^2.

* $R_{merge} = \sum_i \sum_i |I_i - \langle I \rangle| / \sum |\langle I \rangle|$, where I_i are intensity measurements for a reflection, and $\langle I \rangle$ is the mean intensity for this reflection.

† $R_{deriv} = \sum |F_{PH} - F_P| / \sum |F_P|$, where F_{PH} is the structure factor amplitude of the derivative crystal and F_P is that of the native.

‡ $R_{Cullis} = \sum ||F_{PH} \pm F_P| - F_H(\text{calc})| / \sum |F_{PH} - F_P|$, where F_P and F_{PH} are defined as for R_{deriv}, and F_H(calc) is the calculated heavy-atom structure factor amplitud summed over centric data only.

§ Phasing power = $\langle F_H \rangle / E$, the r.m.s. heavy-atom structure factor amplitudes divided by the residual lack of closure error.

bulky end of the molecule. Through their contact one of the two β sheets in domain III is almost entirely buried. To our knowledge (see, for example, ref. 25), the packing of helices in domain I and of sheets in domain II are both novel arrangements.

Domain I. The central helix in this seven-helix bundle is α_5 (Fig. 3*b*,*c*), which is oriented with its C terminus towards the bulky end of the molecule. Viewed from this end, the outer helices are arranged anticlockwise in the order of α_1, α_2, α_3, α_4, α_6 and α_7, with helices α_1 and α_7 adjacent to the β-sheet domains; α_2 is interrupted by a non-helical section and only the leading half, α_{2a}, is packed against α_5. Figure 3*a* shows the alignment of amino-acid sequence on the surfaces of the helices. The helices are long, especially α_3 to α_7, which contain respectively 8, 7, 6, 9 and 7 complete helical turns and hence would be long enough to span the 30-Å thick hydrophobic region of a membrane bilayer. Furthermore, the six outer helices bear a strip of hydrophobic residues (defined by $\Delta G \geq 0$ for transfer from oil to water) down their entire length on the side-facing helix α_5, so they are amphipathic. In keeping with the general observation that secondary structures are close-packed and bury hydrophobic surfaces[26], the helix contact angles in this domain cluster around +20° rather than −50°, giving the bundle a bouquet-like appearance (Fig. 3*b*). Figure 3*c* shows the bundle in cross-section. The interhelical space contains 27 aromatic residues which are packed in the edge-to-face fashion[27]; all polar groups in this region are hydrogen-bonded or in salt bridges.

The concentric arrangement of the seven-helix bundle is dis tinct from the two-layered type seen in bacteriorhodopsin. Ther is some resemblance to the pore-forming domain of colicin A[28] in which two hydrophobic helices are shielded from solvent b eight amphiphilic helices, but the colicin helices are generall shorter. Like the colicin helices, the bundle in the beetle toxi may be a soluble form of packaging for the hydrophobic an amphiphilic helices that will form pores in the membrane afte a large change in conformation.

Domain II. In Fig. 4*a* and 4*b* the three sheets of this domain ar laid side-by-side, as they would be seen from the solvent. Ther is an apparent structural duplication between the four-strande antiparallel sheets, sheet 1 and sheet 2. The chain connections β_4, β_3, β_2, β_5 and β_8, β_7, β_6, β_9, respectively, follow the orde of +3, −1, −1, +3, which is typical of the 'Greek-key' topology[29] From both sheets the inner strands, β_3 and β_2 as well as β_7 an β_6, extend some 20 Å to the apex of the molecule as two-stranded β ribbons; and at the point of departure from the sheets there is a β-bulge in β_3 and in β_7 to twist the plane of the ribbon by nearly 90° relative to the sheet. The connections between the outer strands cross over the ribbons on the solvent side.

The pseudo-symmetry between these sheets is very approximate. Using the least squares option in *O* (ref. 23), the sheet region of the strands β_3 and β_2 can be brought to superimpose on that of β_7 and β_6, with a r.m.s. fit of 0.72 Å for 13 α carbons. But the r.m.s. fit increased to 1.1 Å for 23 α carbons of the

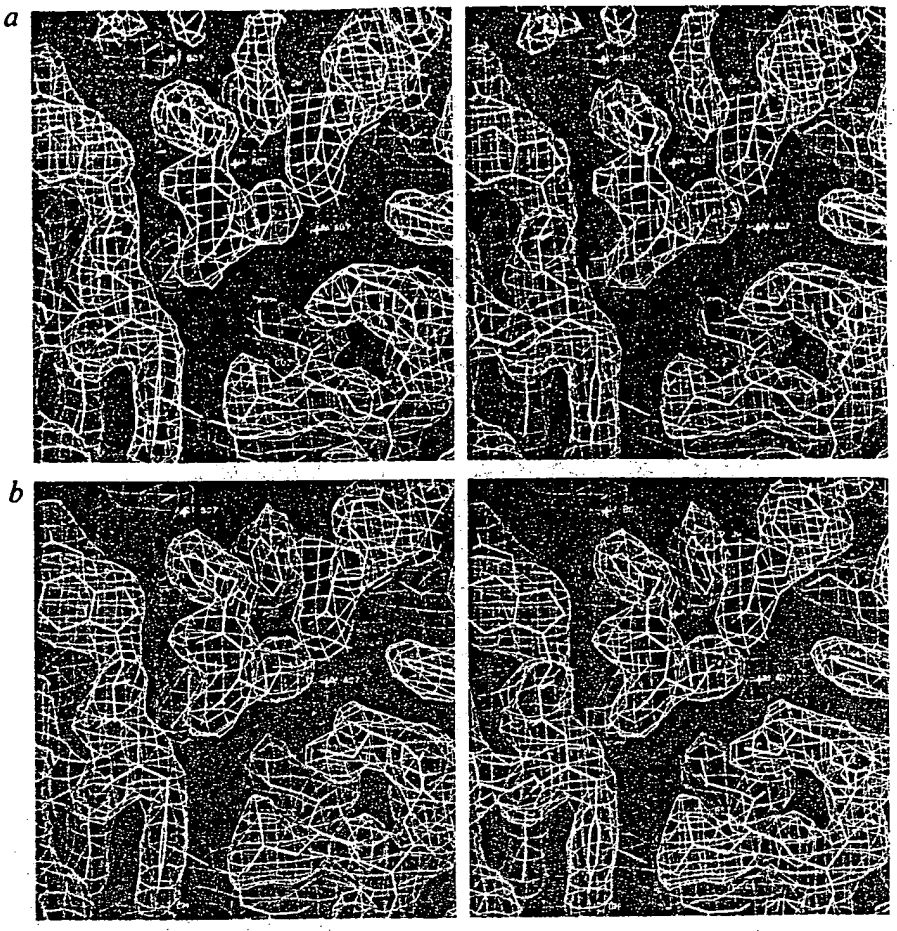


FIG. 1 Electron density map in the neighbourhood of Cys 243, calculated *a*, using combined phases[46] from multiple isomorphous replacement and solvent flattening, and *b*, using combined experimental and model phases[46] after refinement by X-PLOR. The refined structure is shown superimposed for reference. Although Cys 243 is a major site of both the methylmercury (MM) and mercuric acetate (MA) derivatives, the methyl mercury site is in a hydrophobic environment compared with the mercuric acetate site.

whole inner strands including the ribbon region, and 1.7 Å for 36 α carbons on all four strands. Nonetheless, the sequence alignment brought by this superposition of the two sheets revealed a low level of internal homology, with seven pairs of equivalent residues (shown in bold) out of 41 aligned α carbons:

```
338  HRIQFHTRFQP(6)SFNYWS(1)NYVSTRPSI(0)GSNDIITSPF(10)NLEFN  395
402  AVANTNLAVWP(0)SAVYSG(1)TKVEFSQYN(3)DEASTQTYDS(7 )SWDSI  453
```

The three-stranded sheet 3 is formed by two separate polypeptide segments. The C-terminal segment of domain II contributes the two-stranded ribbon of β_{10} and β_{11}, whereas the N-terminal segment of this domain contributes strand β_1, which is hydrogen-bonded to β_{11}; β_1 is followed by a two-turn helix α_8 and an extended chain.

Figure 4*c* and *d* shows in side view and in cross-section that the three antiparallel sheets are packed around a triangular hydrophobic core. This brings the strand β_{10} on the edge of sheet 3 into proximity with strand β_4 on the edge of sheet 1, as well as placing the loops at the end of the three β ribbons into a region of about 12 Å radius at the molecular apex. This domain is in contact with helix α_7 of domain I on the face of sheet 3 (Fig. 4*c*).

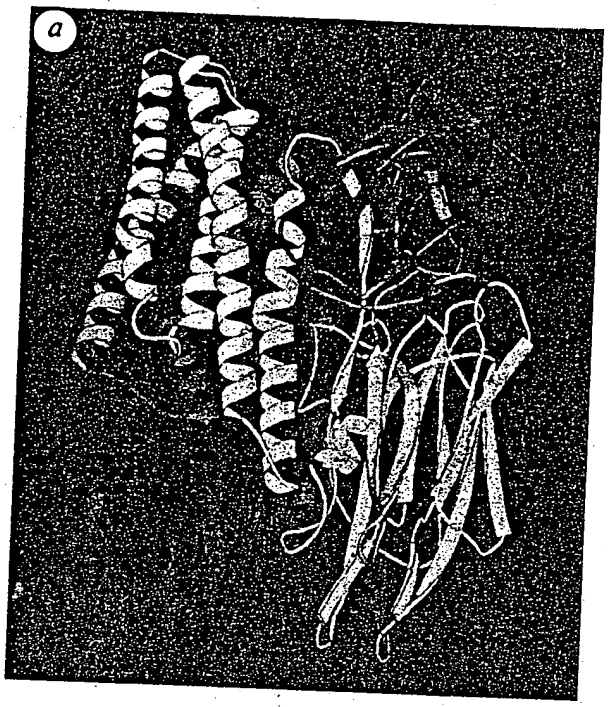
Domain III. Figure 5 is a ribbon drawing of the strands forming the two sheets of the β sandwich. The sheet containing the C-terminal strand is in contact with domain I and will be called the inner sheet. This domain has the 'jelly-roll' topology[29], because it can be generated by folding an antiparallel β ribbon which starts with β_{13} (N terminus) and β_{23} (C terminus) on the inner sheet, and ends in the loop between β_{18} and β_{19} on the outer sheet; β_{14} is a short excursion from this ribbon and forms the fifth antiparallel strand of the outer sheet. In addition, small parallel sheets are formed at the edge of the β sandwich through hydrogen bonding of strand β_{12} to β_{16} at the edge of the outer sheet, and β_1 to β_{13} at the edge of the inner sheet.

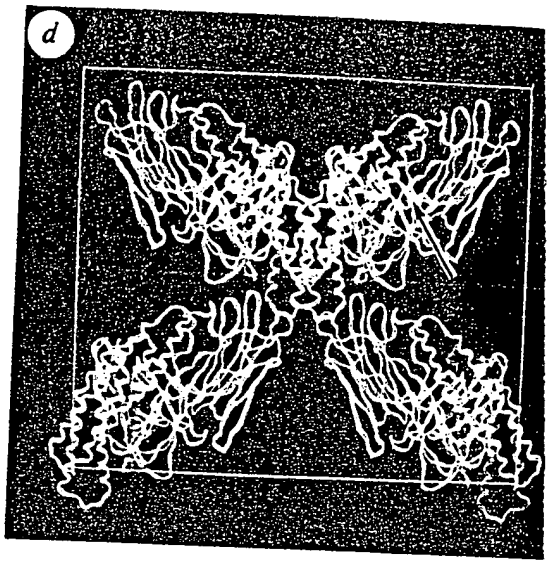
Distribution of conserved sequences. The core of the beetle toxin molecule encompassing the domain interfaces is built from the five sequence blocks that are highly conserved throughout the δ-endotoxin family[1] (Fig. 2*b*,*c*). Block 1, located in the beetle toxin sequence at residues 189–218, corresponds to the central helix (α_5) of the bundle in domain I. Block 2, residues 239–305, overlaps with the latter half of α_6, and with α_7 and β_1; the latter hydrogen-bonds to the edge of the inner sheet in domain III before forming part of the three-stranded sheet 3 in domain II. Block 3, residues 491–538, overlaps with the latter part of β_{11}, where it is hydrogen-bonded to β_1, and with the loops connecting domains II and III. The remainder of block 3 together with blocks 4 and 5, namely residues 560–569 and 633 to the C terminus, respectively, constitute the three buried strands of the inner antiparallel sheet in domain III. The high degree of conservation of internal residues implies that homologous proteins would adopt a similar fold. Using the beetle toxin structure as a model, we can therefore propose a basis for the insecticidal activity of δ-endotoxins as a family.

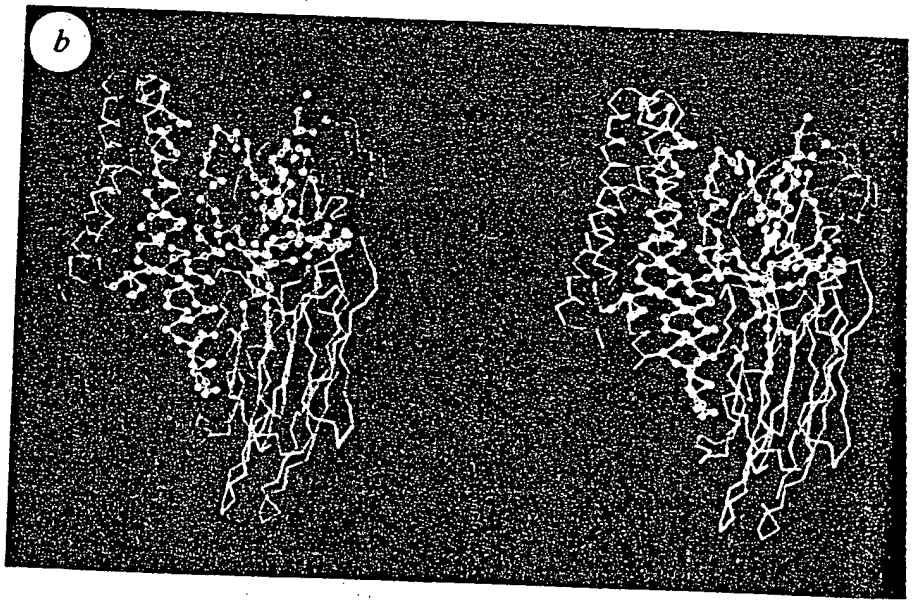
Basis of insecticidal function

Solubility. The beetle toxin crystals are isomorphous with the parasporal crystals[18,19] and show the molecular contacts responsible for solubility behaviour *in vivo*. Four intermolecular salt bridges, Asp 142–Arg 165, Asp 224–Arg 562, Asp 590–Arg 178, and Glu 223–Lys 293, are located at contacts to three different neighbouring molecules. Such salt bridges keep the protoxin crystals insoluble until exposed to the extreme pHs in the insect midgut.

Proteolytic activation. Pro-δ-endotoxins have M_rs of either ~130K or ~70K. Activation by larval gut proteases removes the C-terminal half of the larger protoxins[30,31] and cleaves them at residue 28 or 29 from the N terminus. The smaller protoxins, such as that of the beetle toxin, are processed only at the N







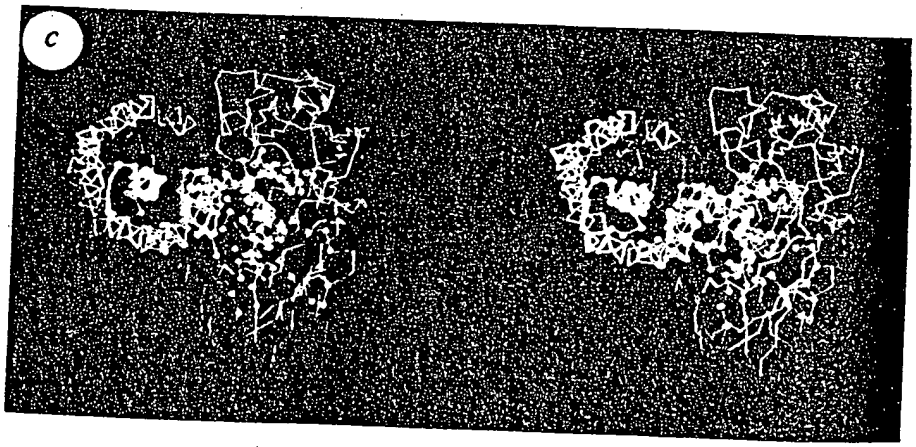


FIG. 2 Overview. *a*, Schematic ribbon representation of the beetle toxin showing the domain organization. Secondary structure assignments are given by Yasspa within program *O* (ref. 23). The polypeptide pathway is indicated by colouring the chain in the rainbow order, from red at the N terminus to blue at the C terminus. The three domains are: I, a seven-helix bundle (upper left); II, a three-sheet assembly (bottom); and III, a β sandwich (upper right). This and all following illustrations of the structure are made with the program MOLSCRIPT[47]. *b* and *c*, Cα trace (stereoview) of the molecule with the five conserved sequence blocks indicated by small beads at their Cα positions. In *b* the view is as in *a*, and in *c* it is down the central helix of the bundle from the bulky end of the molecule; *c* shows that the central helix of domain I and the inner sheet of domain III are conserved; *b* shows that the helices at the domain I–II interface and the loops at the domain II–III interface are also conserved. Note in *c* the helix packing of six around one in domain I. *d*, The solvent channel in the $C222_1$ lattice viewed along the *c* axis. One half of the unit cell thickness is shown, containing four molecules. The other half of the cell is related to this by a two-fold rotation about horizontal axes (blue lines) at $(\frac{1}{2}, y, \pm\frac{1}{4})$. The stacking of both layers leaves solvent channels that traverse the cell along the *c* direction. The N terminus of the molecule (arrow) is accessible from these channels.

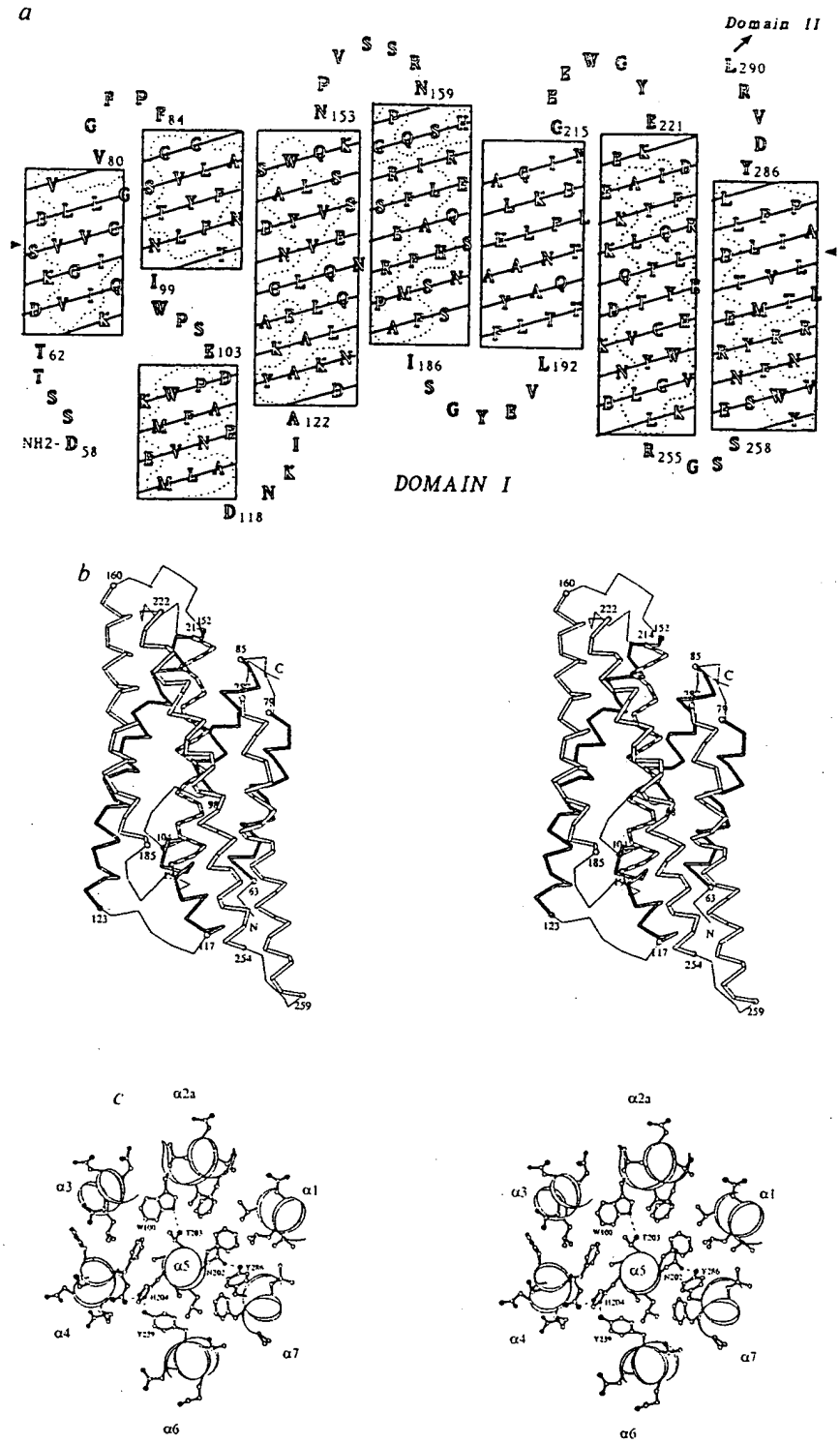


FIG. 3 The seven-helix bundle. *a*, Helical nets showing the position of amino-acid residues along the 7 helices: α_1 (63-79); α_2 (α_{2a}, 85-98 and α_{2b}, 104-117), α_3 (123-152), α_4 (160-185), α_5 (193-214), α_6 (222-254) and α_7 (259-285). The cylindrical surface of the helices are cut longitudinally on the side facing the solvent and flattened to give a view from the interior of the bundle. The top of the drawing corresponds to the bulky end of the whole molecule. Owing to tilting of the outer helices, different helices are in register vertically only at a level indicated by two arrows pointed at α_1 and α_7; α_5 is the central helix. Dotted curves outline the strip of hydrophobic residues down the inward surface of the other six helices. *b*, Cα trace (stereoview) for the bundle viewed perpendicular to α_5. The relative tilt of the outer helices to α_5 and that between adjacent outer heleices are both about 20°. The Cα trace is shaded grey over helices α1 to α3 in the back, striped over helix α5 in the centre, and white over helices α4, α5, and α7 in the front. *c*, Cross-section of the bundle at the level indicated by the arrows in *a*, viewed from the bulky end of the molecule. The hellical backbone is represented by curly ribbons passing through the Cα positions. The outer helices are positioned roughly hexagonally around the central one and tilted relative to it, so the bundle forms a left-handed superhelix. The aromatic side chains are packed in an edge-to-face fashion. Hydrogen bonds are shown for side-chain atoms.

terminus[19,32] where about 50 residues are removed. The activated δ-endotoxins show a conserved C-terminus, so-called sequence block 5 (ref. 1). Its position as the middle strand of the buried β sheet in domain III precludes further processing from the C terminus. In fact deletion from this site by 4 to 8 residues results in inactive mutants with altered solubility and immunogenicity[30,33-35]. This is not surprising as the inner sheet can be expected to play a critical part in the structural integrity and stability of the toxins through interaction with the helical bundle.

At the N-terminal cleavage sites the different protoxin sequences show locally similar hydropathy profiles[36,37], which would be consistent with a common topology for the N-terminal region of the activated toxins as seen in the helical bundle of the beetle toxin. In crystals of the beetle toxin, the N terminus at the start of helix α_1 borders on a large solvent channel of about 30 Å diameter that crosses the unit cell along the *c* direction (Fig. 2*d*). This channel could allow access of sporulation-associated proteases to the cleavage site in parasporal crystals[19].

Receptor binding. The insecticidal selectivity of δ-endotoxins is due to high-affinity binding to specific membrane receptors[7-9,38], which in three cases seem to be glycoproteins[38-40]. For several δ-endotoxins the specificity-determining regions have been delimited by exchanging sequence segments between closely related toxins of differing specificities[13-15]. Guided by the location of secondary structures in the beetle toxin, a plausible alignment of δ-endotoxin sequences was made for the non-

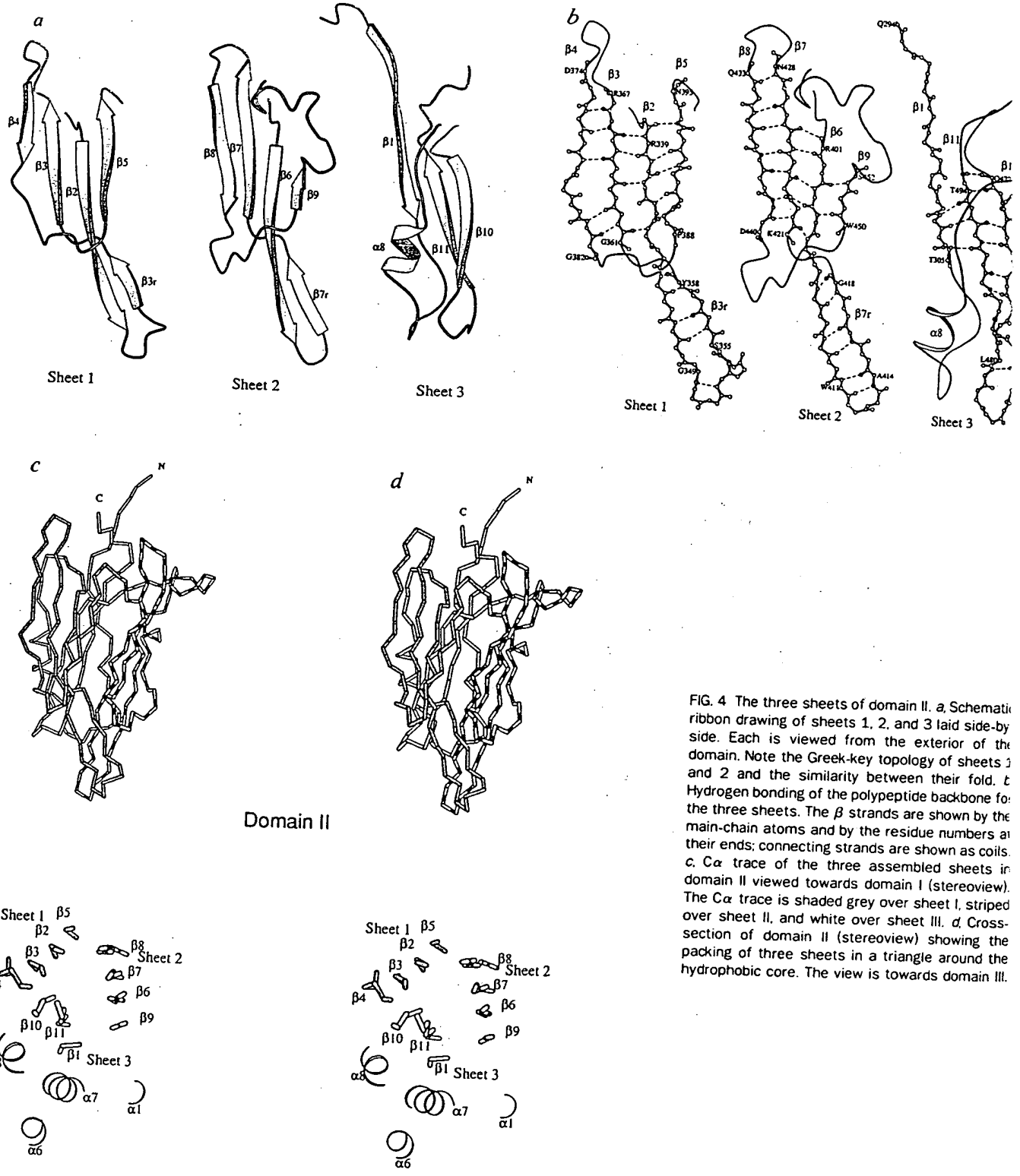


FIG. 4 The three sheets of domain II. *a*, Schematic ribbon drawing of sheets 1, 2, and 3 laid side-by-side. Each is viewed from the exterior of the domain. Note the Greek-key topology of sheets 1 and 2 and the similarity between their fold. *b*, Hydrogen bonding of the polypeptide backbone for the three sheets. The β strands are shown by the main-chain atoms and by the residue numbers at their ends; connecting strands are shown as coils. *c*, Cα trace of the three assembled sheets in domain II viewed towards domain I (stereoview). The Cα trace is shaded grey over sheet I, striped over sheet II, and white over sheet III. *d*, Cross-section of domain II (stereoview) showing the packing of three sheets in a triangle around the hydrophobic core. The view is towards domain III.

conserved regions (ref. 12, and T. C. Hodgman, unpublished results). Hence the genetically identified specificity-determining regions can be mapped to equivalent positions in the beetle toxin structure, and these fall mainly in domain II. For instance, the dual specificity of CryIIA for Lepidoptera and Diptera, as distinct from the Lepidoptera specificity in the closely related CryIIB, is determined by residues 307–382 of their sequences[14], which corresponds roughly to sheet 1 (Fig. 4*a*) plus strand β_6 in sheet 2 and the loop leading up to β_7, whereas the Lepidoptera specificity of CryIIB is dependent on a longer segment[14] that would include both inner strands of sheet 2. Similarly, the toxicities of CryIA(a) and CryIA(c) to two lepidopteran insects depend on three segments termed x, y and z (ref. 15): amino-acid substitutions in y can reduce toxicity by up to 2,000-fold, and segments x and y interact in determining specificity. Aligned with the beetle toxin structure, segment x corresponds roughly to the outer strands β_4 and β_5 of sheet 1 and the whole of sheet 2, including the loop entering β_{10} in sheet 3; y corresponds to

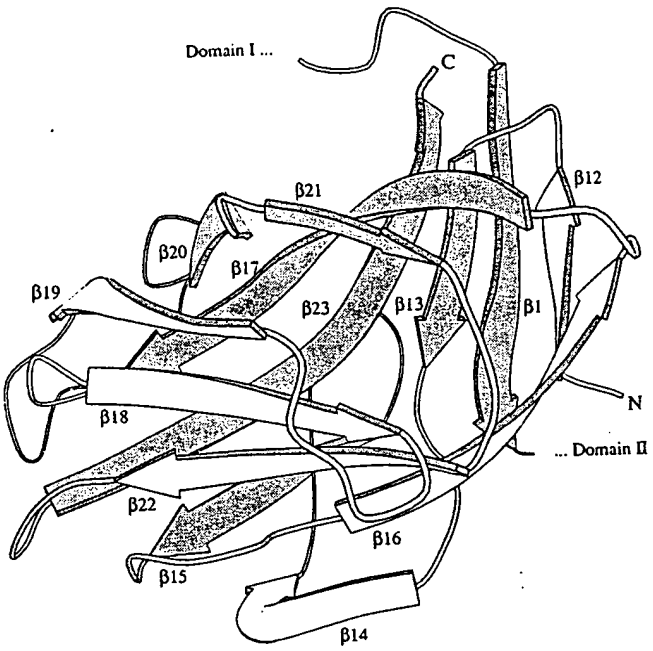


FIG. 5 Domain III, schematic ribbon representation of the β sandwich. β strands forming the inner sheet are shaded grey. The topology of an eight-stranded 'jelly-roll' can be seen by following the β hairpin starting with β_{13}, β_{15} and β_{23} in the inner sheet, continuing to β_{16} and β_{22} in the outer sheet, then β_{17} and β_{21}, β_{20} in the inner sheet, and ending with β_{18} and β_{19} in the outer sheet. β_{14} is an excursion from the hairpin and forms a fifth antiparallel strand of the outer sheet. Small parallel β sheets are added to one edge of the β sandwich, by hydrogen bonding of β_1 to β_{13} in the inner sheet and β_{12} to β_{16} in the outer sheet. Residue numbers in the β strands are: β_{12}, 502-506; β_{13}, 509-513; β_{14}, 519-525; β_{15}, 536-541; β_{16}, 547-554; β_{17}, 558-569; β_{18}, 573-579; β_{19}, 585-591; β_{20}, 604-609; β_{21}, 611-614; β_{22}, 619-625; and β_{23}, 631-643.

strand β_{10} of sheet 3 and the loop connecting β_{10} and β_{11}; and z extends from β_{11} to the C-terminal activation site. Furthermore, the interaction between x and y can be understood in terms of the proximity between β_4 on the edge of sheet 1 and β_{10} on the edge of sheet 3. Although z was inferred[15] to extend into domain III, the combined evidence from genetics and receptor-binding assays *in vitro* for Lepidoptera toxins[9,41] correlates receptor recognition with sequence variations within domain II. We note that the β ribbons from all three sheets terminate in loops in a small region on the molecular apex, in a manner reminiscent of the complementarity-determining region of immunoglobins.

Pore formation. The common mechanism of epithelial cell disruption by δ-endotoxins of widely different specificities is believed to be the formation of lytic pores of 10 to 20 Å diameter in the insect membrane[10]. The structure of the beetle toxin displays an apparatus for pore formation in the long, hydrophobic and amphipathic helices of domain I which could penetrate the membrane. Between the crystal structure in which the bouquet-like helical bundle internalizes all the hydrophobic surfaces, and the unknown pore structure where hydrophobic surfaces would be in intimate contact with the membrane lipids, large conformation changes must occur. In the absence of a full characterization of the pore-forming process, we propose the following by extrapolation from the crystal structure.

The trigger for the conformational changes may be provided by receptor binding and the consequent interaction of toxin with the membrane bilayer. Membrane insertion follows rapidly, so that a major part of the bound δ-endotoxin cannot be displaced from the brush-border vesicles by other toxins recognizing the same receptor sites[7,9]. As domain II and probably its apical region are most likely to bind the membrane receptors, the helices are expected to insert with the 'domain II end' (see Fig. 2*a*) oriented towards the cytoplasm. If helical hairpins are to initiate the membrane penetration, as probably happens for colicin[28,42,43], they will probably be linked at the domain II end. So either of the helix pairs α_6-α_7 or α_4-α_5 could be the likely initiator. The α_6-α_7 pair is favoured because it forms part of the conserved interface with domain II and is well positioned to sense the receptor binding. On the other hand, helix α_5 is the most conserved throughout the family of δ-endotoxins. Point mutations in α_5 reduce toxicity of a Lepidoptera toxin without reducing binding to membranes[44]. Proteolysis in the interhelical loops at the domain III end, as in the α_3-α_4 loop[19,32], may facilitate release of the helix pairs from the tertiary structure of the bundle. The insertion of a hairpin can create a defect in the membrane, allowing the rest of domain I to participate in pore formation in a cooperative manner. □

Received 22 July; accepted 19 September 1991.

1. Höfte, H. & Whiteley, H. R., *Microbiol. Rev.* **53**, 242-255 (1989).
2. Ellar, D. J. *et al.* in *Molecular Biology of Microbial Differentiation* (eds Hoch, J. A. & Setlow, P.) 230-240 (Am. Soc. Microbiol., Washington, DC, 1985).
3. Wilcox, E. R. *et al.* in *Protein Engineering: Applications in Science, Medicine and Industry* (eds Inouye, M. & Sarma, R.) 395-413 (Academic, New York, 1986).
4. Vaeck, M. *et al. Nature* **328**, 33-37 (1987).
5. Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, D. L. & Fischhoff, D. A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3324-3328 (1991).
6. Barton, K. A., Whiteley, H. R. & Yang, N. S. *Plant Physiol.* **85**, 1103-1109 (1987).
7. Hofmann, C., Lüthy, P., Hütter, R. & Pliska, V. *Eur. J. Biochem.* **173**, 85-91 (1988).
8. Hofmann, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7844-7848 (1988).
9. Van Rie, J., Jansens, S., Höfte, H., Degheele, D. & Van Mellaert, H. *Appl. Envir. Microbiol.* **56**, 1378-1385 (1990).
10. Knowles, B. H. & Ellar, D. J. *Biochim. biophys. Acta* **924**, 509-518 (1987).
11. Endo, Y. & Nishitsutsuji-Uwo, J. *J. Invertebr. Path.* **36**, 90-103 (1980).
12. Hodgman, T. C. & Ellar, D. J. *J. DNA Sequ. Map.* **1**, 97-106 (1990).
13. Ge, A. Z., Shivarova, N. I. & Dean, D. H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4037-4041 (1989).
14. Widner, W. R. & Whiteley, H. R. *J. Bact.* **172**, 2826-2832 (1990).
15. Schnepf, H. E., Tomczak, K., Ortega, J. P. & Whiteley, H. R. *J. biol. Chem.* **265**, 20923-20939 (1990).
16. Krieg, A., Huger, A. M., Langenbruch, G. A. & Schnetter, W. *J. appl. Entomol.* **96**, 500-508 (1983).
17. Höfte, H., Seurinck, J., Van Houtven, A. & Vaeck, M. *Nucleic Acids Res.* **15**, 7183 (1987).
18. Li, J., Henderson, R., Carroll, J. & Ellar, D. *J. molec. Biol.* **199**, 543-545 (1988).
19. Carroll, J., Li, J. & Ellar, D. J. *Biochem. J.* **261**, 99-105 (1989).
20. Arndt, U. W. *Meth. Enzym.* **114**, 472-485 (1985).
21. Wang, B. C. *Meth. Enzym.* **115**, 90-112 (1985).
22. Leslie, A. G. W. *Acta crystallogr.* **A43**, 134-136 (1987).
23. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, N. *Acta crystallogr.* **A47**, 110-119 (1991).
24. Brünger, A. T. *J. molec. Biol.* **203**, 803-816 (1988).
25. Chothia, C. & Finkelstein, A. V. *A. Rev. Biochem.* **59**, 1007-1039 (1990).
26. Chothia, C. *J. molec. Biol.* **105**, 1-15 (1976).
27. Burley, S. K. & Petsko, G. A. *Science* **229**, 23-28 (1985).
28. Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. *Nature* **337**, 93-96 (1989).
29. Richardson, J. S. *Adv. Prot. Chem.* **34**, 167-339 (1981).
30. Höfte, H. *et al. Eur. J. Biochem.* **161**, 273-280 (1986).
31. Choma, C. T. & Kaplan, H. *Biochemistry* **29**, 10971-10977 (1990).
32. Nicholls, C. N., Ahmad, W. & Ellar, D. J. *J. Bact.* **171**, 5141-5147 (1989).
33. MacIntosh, S. C., McPherson, S. L., Perlak, F. J., Marrone, P. G. & Fuchs, R. L. *Biochem. biophys. Res. Commun.* **170**, 665-672 (1990).
34. Schnepf, H. E. & Whiteley, H. R. *J. biol. Chem.* **260**, 6273-6280 (1985).
35. Adang, M. J. *et al. Gene* **36**, 289-300 (1985).
36. Sekar, R., Thompson, D. V., Maroney, M. J., Bookland, R. G. & Adang, M. J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7036-7040 (1987).
37. Chungjatupornchai, W., Höfte, H., Seurinck, J., Angsuthanasombat, C. & Vaeck, M. *Eur. J. Biochem.* **173**, 9-16 (1988).
38. Knowles, B. H., Thomas, W. E. & Ellar, D. J. *FEBS Lett.* **168**, 197-202 (1984).
39. Knowles, B. H. & Ellar, D. J. *J. Cell Sci.* **83**, 89-101 (1986).
40. Haider, M. Z. & Ellar, D. J. *Biochem. J.* **248**, 197-201 (1987).
41. Visser, B., Munsterman, E., Stoker, A. & Dirske, W. G. *J. Bact.* **172**, 6783-6788 (1990).
42. Lakey, J. H., Baty, D. & Pattus, F. *J. molec. Biol.* **218**, 639-653 (1991).
43. Song, H. Y., Cohen, F. S. & Cramer, W. A. *J. Bact.* **173**, 2927-2934 (1991).
44. Ahmad, W. & Ellar, D. J. *FEMS Microbiol. Lett.* **68**, 97-104 (1990).
45. Messerschmidt, A. & Pflugrath, J. W. *J. appl. Crystallogr.* **20**, 306-315 (1987).
46. Read, R. J. *Acta crystallogr.* **A42**, 140-149 (1986).
47. Kraulis, P. J. *appl. Crystallogr.* (in the press).

ACKNOWLEDGEMENTS. We thank P. R. Evans, A. G. W. Leslie and R. Henderson for advice and encouragement; K. Wilson and Z. Dauter for help with the image plate system; K. Nagai and P. J. McLaughlin for help in collecting film data; T. A. Jones for advice on model building; SERC Daresbury Laboratory and EMBL Outstation at DESY for use of synchrotron facilities; and T. Woollard and K. Hopkins for maintaining the rotating anodes. D.J.E. and J.C. acknowledge the support of the AFRC.

APPENDIX B

Pfam 16.0 (Saint Louis)
Home | Analyze a sequence | Browse Pfam | Keyword search | Taxonomy search | Swisspfam | PfamAlyzer | Help

There are 7 searches queued ahead of you on the Pfam compute server.
Please wait...

Starting search. Estimated time: 38 seconds (assuming all Wulfpack nodes are running). Please wait...

Pfam HMM search results, glocal+local alignments merged (Pfam_ls+Pfam_fs)

[Go here for an explanation of the format of the results]

	Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value	Alignment	Description
!!	Endotoxin_N	70	293	1	244	457.8	**1.2e-134**	glocal	delta endotoxin, N-terminal domain
!!	Endotoxin_M	298	513	1	242	204.3	**2.5e-58**	glocal	delta endotoxin
!!	Endotoxin_C	523	663	1	155	121.1	**2.6e-33**	glocal	delta endotoxin

Alignments of top-scoring domains:

Format for alignment of query to Seed: Plain text

```
Endotoxin_N: domain 1 of 1, from 70 to 293: score 457.8, E = 1.2e-134
                  *->vqiglsivgtlLgalGvfPggGflvgfystLldlLWPsngpsnenvW
                     ++++++ivg+lL+ lGv P++G +v++y++L+d+LWPs+++s   +W
       query    70    AKAAIDIVGKLLSGLGV-PFVGPIVSLYTQLIDILWPSGEKS---QW 112

                     eaFleqvEqLIdQrIseyvrnrAiarLeGLgnsydteViYleaLeeWekn
                     e+F+eqvE+LI+Q+I+ey+rn+A+++LeGLgn+y+    Yl aLeeWe+n
       query   113 EIFMEQVEELINQKIAEYARNKALSELEGLGNNYQ---LYLTALEEWEEN 159

                     pnnarsreaVrtrFnildslfvnaipsFavsagysenyevlLLPvYAQAA
                     pn+ r+  +Vr rF+ildslf++ +psF+v+ ++ e   v++L+vYA+AA
       query   160 PNGSRALRDVRNRFEILDSLFTQYMPSFRVT-NF-E---VPFLTVYAMAA 204

                     NLHLlLLRDAvifGerWgltqadinstldednyYnrllerikeYtdHCvn
                     NLHLlLL+DA+ifGe+Wg ++++in      nyY r+ ++++eY+dHCv+
       query   205 NLHLLLLKDASIFGEEWGWSTTTIN------NYYDRQMKLTAEYSDHCVK 248
```

```
                wYNtGLnnlrgtnldaesWvrYNryRReMTLtVLDlVAlFPnYDprl<-*
                wY tGL++l+gt+  a++Wv+YN++RReMTL VLD VAlFPnYD+r+
query    249 WYETGLAKLKGTS--AKQWVDYNQFRREMTLAVLDVVALFPNYDTRT    293
```

align query/70-293 to Endotoxin_N(ls) Seed

```
Endotoxin_M: domain 1 of 1, from 298 to 513: score 204.3, E = 2.5e-58
                *->tksqLTREiYTDPvgevspgsglseglcrrWginnyprltFsalEna
                   tk+qLTRE+YTDP+g v+        ++  ++++ +p  +F ++E++
query    298    TKAQLTREVYTDPLGAVN-------VSSIGSWYDKAP--SFGVIESS 335

                liRsPHLfdfLnsltiyTnssrgplnttldinyWsGhrvtssytggstln
                 iR+PH fd++  lt yT s  ++++  + i++W+Gh+++++++
query    336 VIRPPHVFDYITGLTVYTQSR-SISSARY-IRHWAGHQISYHRVSR---- 379

                niissplyGnttntaeppvtispcftnndiYRtlsatsnrlsgnniigln
                +++    yG++ n +  ++t+   ftn diY+tls + ++l+    +g++
query    380 GSNLQQMYGTNQN-LHSTSTFD--FTNYDIYKTLSKDAVLLD-IVYPGYT 425

                npingvtrvdFygangtnseissntyrss.krgnggqrtidsideLPpet
                 + +g + v+F++ n  n+++ + +y++ +k  +    t+ds+ eLPpet
query    426 YIFFGMPEVEFFMVNQLNNTRKTLKYNPVsKDIIAS--TRDSELELPPET 473

                tnePiyesYSHrLShvtflrsnttqggsdatrahvpvFsWTHrSad<-*
                +++P+yesYSHrL+h+t++  + +  g       vpvFsWTHrSad
query    474 SDQPNYESYSHRLCHITSIPATGNTTGL------VPVFSWTHRSAD    513
```

align query/298-513 to Endotoxin_M(ls) Seed

```
Endotoxin_C: domain 1 of 1, from 523 to 663: score 121.1, E = 2.6e-33
                *->ITQIPlVKaynlssgasVVkGPGFTGGDilrrtssnGsfgtlrvttk
                   ITQIP+VK         VVkGPG+TGGD+l+ + s+Gs+gtl +
query    523    ITQIPAVKCWDNLPFVPVVKGPGHTGGDLLQYNRSTGSVGTLFLARY 569

                linnplsqrYRiRIRYASttnlrfivsliggttsnqfnfpkTmnrgdnye
                  ++    +YR+R+RYA+ ++++++v+       +q+  pkTmn g   e
query    570 GLALEKAGKYRVRLRYATDADIVLHVN------DAQIQMPKTMNPG---E 610

                dLtYesFryaefstpvfspyfsgsqdiltnistlgiqgfssggnqevYID
                dLt+++F+ a+  t+ ++    ++++ l +  +lg ++ s+  ++ vY+D
query    611 DLTSKTFKVADAITT-LN---LATDSSLALKHNLGEDPNSTL-SGIVYVD 655

                rIEFIPvn<-*
                rIEFIPv+
query    656 RIEFIPVD    663
```

align query/523-663 to Endotoxin_C(ls) Seed

NEW! Phylogenomic analysis of query using RIO.

Given a query sequence, Pfam domain, and species, the RIO server will order sequences

in the Pfam domain by orthology to the query. Many other options are available, and an annotated gene tree can be generated and viewed with ATV. The button below will send your query and Pfam domain hits to the RIO server.

send query to RIO

Home | Analyze a sequence | Browse Pfam | Keyword search | Taxonomy search | Swisspfam | PfamAlyzer | Help
Comments, questions, flames? Email <pfam@genetics.wustl.edu>.

Endotoxin_M <-- --> Enhancin

Endotoxin_N

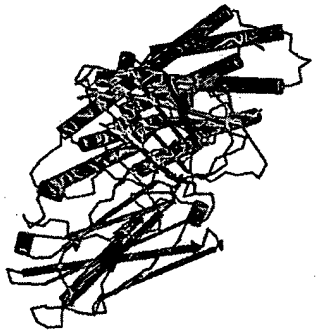


Figure 1: 1dlc Toxin

Image from PDBsum database

***Accession number:* PF03945**
delta endotoxin, N-terminal domain

This family contains insecticidal toxins produced by Bacillus species of bacteria. During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N terminus is cleaved in all of the proteins and a C terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description

This family contains insecticidal toxins produced by Bacillus species of bacteria. During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N-terminus is cleaved in all of the proteins and a C-terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description text from InterPro entry IPR005639

Sequence information

Alignment

◉ Seed (38) ○ Full (173)

Format:

Hyperlinked plain text

Retrieve alignment

Visualize domain structures

◉ Seed (38) ○ Full (173)

display 10 per page.

Retrieve domain structures

Species distribution

Tree depth:

all

View species tree

Literature References

[1]
Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 angstroms resolution.
Li J, Carroll J, Ellar DJ;
Nature 1991;353:815-821.

[2]
Bacillus thuringiensis and its pesticidal crystal proteins.
Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH;
Microbiol Mol Biol Rev 1998;62:775-806.

Database References

HOMSTRAD	endotoxin
PDB	1i5p 1dlc 1ciy 1ji6
SCOP	1dlc (family)
INTERPRO	IPR005639

HMMER build information

	Pfam_ls [download HMM]	Pfam_fs [download HMM]
Gathering cutoff	-55.00 -55.00	10.00 10.00
Trusted cutoff	-52.10 -52.10	10.00 10.00
Noise cutoff	-73.30 -73.30	9.50 9.90
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

Pfam specific information

Author of entry	Bateman A, de Maagd R
Type definition	Domain
Alignment method of seed	
Source of seed members	Arne Eloffson

Endotoxin_C <-- --> Endotoxin_N

Endotoxin_M

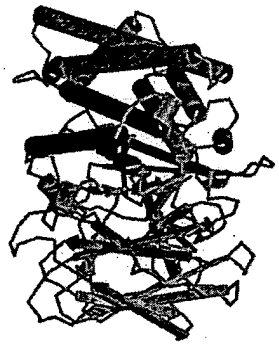


Figure 1: 1ciy
Toxin
Insecticidal toxin: structure and channel formation

Image from PDBsum database

Accession number: **PF00555**
delta endotoxin

This family contains insecticidal toxins produced by Bacillus species of bacteria. During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N terminus is cleaved in all of the proteins and a C terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description

This entry contains insecticidal toxins produced by Bacillus species of bacteria.During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N terminus is cleaved in all of the proteins and a C-terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description text from InterPro entry IPR001178

Sequence information

Alignment

◉ Seed (38) ○ Full (140)

Format:

Visualize domain structures

◉ Seed (38) ○ Full (140)

Species distribution

Tree depth:

Hyperlinked plain text | display 10 per page. | all

Retrieve alignment | Retrieve domain structures | View species tree

Literature References

[1]
Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 angstroms resolution.
Li J, Carroll J, Ellar DJ;
Nature 1991;353:815-821.

[2]
Bacillus thuringiensis and its pesticidal crystal proteins.
Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH;
Microbiol Mol Biol Rev 1998;62:775-806.

Database References

HOMSTRAD	endotoxin
PDB	1ciy 1dlc 1ji6
PFAMB *The following Pfam-B family may contain sequences that according to Prodom are members of this Pfam-A family.*	PB054837
SCOP	1dlc (family)
INTERPRO	IPR001178

HMMER build information

	Pfam_ls [download HMM]	Pfam_fs [download HMM]
Gathering cutoff	-30.00 -30.00	10.00 10.00
Trusted cutoff	-26.00 -26.00	10.60 10.60
Noise cutoff	-36.80 -36.80	8.80 9.40
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

Pfam specific information

Author of entry	Bateman A, de Maagd R
Type definition	Domain

Alignment method of seed

Source of seed members Arne Eloffson

Home | Analyze a sequence | Browse Pfam | Keyword search | Taxonomy search | Swisspfam | PfamAlyzer | Help

Comments, questions, flames? Email <pfam@genetics.wustl.edu>.

Endothelin <-- --> Endotoxin_M

Endotoxin_C

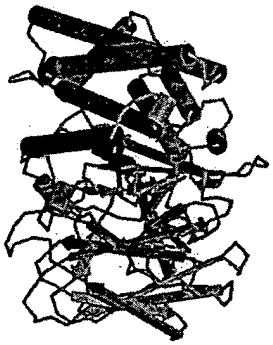


Figure 1: 1ciy
Toxin
Insecticidal toxin: structure and channel formation

Image from PDBsum database

Accession number: **PF03944**
delta endotoxin

This family contains insecticidal toxins produced by Bacillus species of bacteria. During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N terminus is cleaved in all of the proteins and a C terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description

This family contains insecticidal toxins produced by Bacillus species of bacteria. During spore formation the bacteria produce crystals of this protein. When an insect ingests these proteins they are activated by proteolytic cleavage. The N-terminus is cleaved in all of the proteins and a C-terminal extension is cleaved in some members. Once activated the endotoxin binds to the gut epithelium and causes cell lysis leading to death. This activated region of the delta endotoxin is composed of three structural domains. The N-terminal helical domain is involved in membrane insertion and pore formation. The second and third domains are involved in receptor binding.

Description text from InterPro entry IPR005638

Sequence information

Alignment

◉ Seed (38) ○ Full (159)

Format:

Hyperlinked plain text

Retrieve alignment

Visualize domain structures

◉ Seed (38) ○ Full (159)

display 10 per page.

Retrieve domain structures

Species distribution

Tree depth:

all

View species tree

Literature References

[1]
Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 angstroms resolution.
Li J, Carroll J, Ellar DJ;
Nature 1991;353:815-821.

[2]
Bacillus thuringiensis and its pesticidal crystal proteins.
Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH;
Microbiol Mol Biol Rev 1998;62:775-806.

Database References

HOMSTRAD	endotoxin
PDB	1i5p 1ciy 1dlc 1ji6
SCOP	1dlc (family)
INTERPRO	IPR005638

HMMER build information

	Pfam_ls [download HMM]	Pfam_fs [download HMM]
Gathering cutoff	-35.00 -35.00	13.00 13.00
Trusted cutoff	-34.80 -34.80	13.50 13.50
Noise cutoff	-36.10 -36.10	9.40 9.40
Build method of HMM	hmmbuild -F HMM_ls SEED hmmcalibrate --seed 0 HMM_ls	hmmbuild -f -F HMM_fs SEED hmmcalibrate --seed 0 HMM_fs

Pfam specific information

Author of entry	Bateman A, de Maagd R
Type definition	Domain
Alignment method of seed	
Source of seed members	Arne Eloffson

APPENDIX C

Attorney Docket No. 035718/237005 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re:	Abad *et al.*	Confirmation No.:	5409
Appl. No.:	10/032,717	Group Art Unit:	1638
Filed:	10/23/2001	Examiner:	A.R. Kubelik
For:	GENES ENCODING NOVEL BACILLUS THURINGIENSIS PROTEINS WITH PESTICIDAL ACTIVITY AGAINST COLEOPTERANS		

January 18, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RULE 132 DECLARATION
of
André Abad

Sir:

I, André Abad, do hereby declare and say as follows:

I am skilled in the art of the field of the invention of the above-referenced application. I earned the following academic degrees: BS majoring in mathematics and biochemistry from the Wisconsin University River Falls in 1978 and Ph.D. from Purdue University Department of Agronomy in 1996. My thesis investigated the role of a mitochondrial gene in cytoplasm male sterility in beans. From 1979 to 1991, I was employed by the University of Minnesota, Department of Plant Pathology. Working in Dr. Blanchette's laboratory, we investigated and published numerous manuscripts in the area of woody tissue degradation by fungi, in particular related to the degradation of cell wall component by fungal enzymes such as xylanase. From 1996 to 1998, I worked in Dr. Judy Bond's laboratory at Hershey Medical Center in Pennsylvania. I was involved in characterization of mouse meprin receptors and generated the constructs and ES cells necessary for producing transgenic mice targeting the knockout of meprin. Since 1999, Pioneer Hi-Bred has employed me. My current responsibility as a research scientist is to

lead a research team for insecticidal protein optimization and genomic screening of *Bacillus thuringiensis* DNA for novel insecticidal genes

1. I am familiar with the experiments described in the above-mentioned application. Particularly, the procedures described in Examples 4, 6, and 7 of the above-mentioned application are considered "routine" by scientists who are familiar with research on endotoxins. Moreover, the production of plants expressing proteins having pesticidal activity, while it is a time-consuming and laborious task, is also considered "routine" by scientists who are responsible for producing such plants.

2. As one of skill in the art, given the disclosure in the above-referenced application, I would be able to make and use the claimed invention. For example, I would be able to make and use the nucleic acid of claim 1 by generating a collection of nucleic acids comprising a nucleotide sequence meeting the sequence limitation of the claims (*i.e.*, a nucleotide sequence having at least 90% sequence identity to the nucleotide sequence set forth in SEQ ID NO:1) and assaying the encoded polypeptide for defensive activity as described in the specification, for example, as described in any of Examples 4, 6, and 7. Further, where such a collection included dozens of such sequences, I would consider this degree of experimentation to be routine rather than to be "undue experimentation." For these reasons, I believe that the claims, including for example, claim 1, are fully enabled and described by the specification.

3. It is my understanding, as one of skill in the art, that proteins can be produced that share a relatively low degree of sequence identity—maybe even as low as 70% sequence identity--with a known protein but that have the same or essentially the same function.

4. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 01-19-2005

By: _____

André Abad

APPENDIX D

Protein evolution by molecular breeding

Jeremy Minshull* and Willem PC Stemmer†

Natural evolution has guided the development of 'molecular breeding' processes used in the laboratory for the rapid modification of subgenomic sequences including single genes. The most significant recent development has been the *in vitro* permutation of natural diversity. Homologous recombination of multiple related sequences produced high-quality libraries of chimeric sequences encoding proteins with functions that differ dramatically from any of the parents. Increasingly powerful screening methods are also being developed, allowing these libraries to be screened for novel biocatalysts.

Addresses
Maxygen Incorporated, 515 Galveston Drive, Redwood City, CA 94063, USA
*e-mail: jeremy_minshull@maxygen.com
†e-mail: pim_stemmer@maxygen.com

Current Opinion in Chemical Biology 1999, 3:284–290

http://biomednet.com/elecref/1367593100300284

 ISSN 1367-5931

Introduction

Enzymes are used in a wide variety of applications including food and feed processing, laundry detergents, chemicals production, paper bleaching and pharmaceutical manufacturing. The benefits of using enzymes as catalysts are that reactions can occur at moderate temperatures, toxic solvents or reactants can often be eliminated, and reactions are usually stereospecific, which is of particular benefit in the synthesis of pharmaceuticals and fine chemicals. The specificity of enzymes also obviates the need for protecting and deprotecting reactive groups, which is a source of considerable yield loss in organic syntheses.

Although three billion years of evolution have produced a wealth of protein catalysts, they are generally not optimal for a particular industrial application. While it is possible to screen enzymes from extremophiles for activity under the appropriate process reaction conditions [1,2], natural selection has selected enzymes to function in the complex mixtures of molecules within cells rather than in bioreactors. Obtaining the desired combinations of properties therefore generally requires further protein optimization.

Structural information has been used with some success to improve enzyme function [3–5]. As a general method, however, structure-based methods require time and equipment in order to generate and process very large amounts of information.

An alternative strategy to making defined changes on the basis of structural understanding is to harness the Darwinian power of recursive cycles of mutation and selection. By using directed evolution, protein engineers attempt to mimic the natural processes by which protein variants arise and are tested for 'fitness' within living systems. In this review, we will focus on the underlying rationale behind and recent advances in directed evolution, both in the methods used to generate protein variants, and in the screening strategies used to identify variants of interest.

DNA shuffling

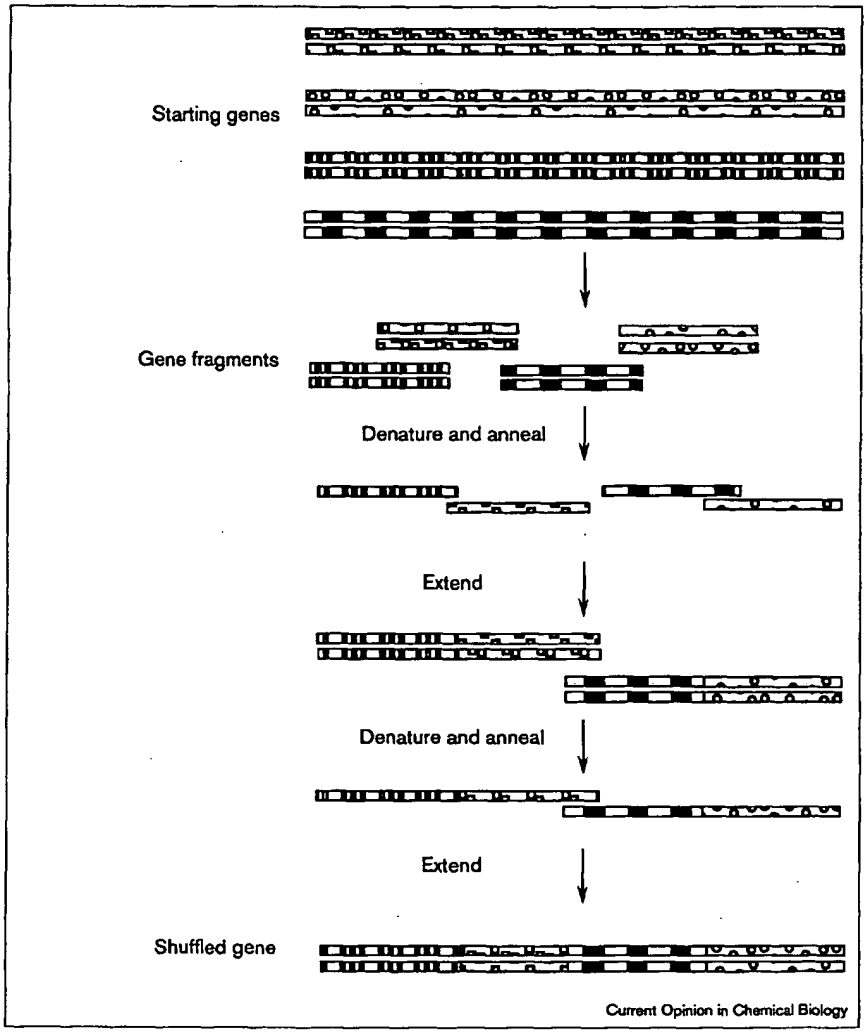
Directed evolution effectively performs the complex computations required to determine the effects of changes in sequence on catalytic function. In addition to the active-site geometry, the impact of sequence changes on protein expression, stability and folding, and interactions with other host proteins and small molecules are all simultaneously considered simply by directly measuring the activity of the mutant enzymes or metabolic pathways.

The best evolutionary strategies are likely to be those that most closely mimic natural ones: in three billion years, not only have individual genes evolved, but the evolutionary process itself has been optimized [6]. Those algorithms that are best at searching through the possible combinations of nucleotides for sequences with biological function have been preserved along with the sequences whose evolution they have facilitated. Recombination is such a mechanism, found universally in biological systems. Genetic algorithms and other computer simulations of simple evolving systems that incorporate the ability to recombine information are more powerful and evolve more rapidly than those which do not [6–9].

Incorporation of recombination into a method for directed evolution of single genes (known as 'DNA shuffling' or 'molecular breeding') was developed recently [10]. In this method, a population of mutant genes (rather than just one) are selected on the basis of their containing beneficial mutations, thus making them appropriate as parents for the next cycle. The genes are randomly fragmented, then reassembled by recombination with each other. The process is shown schematically in Figure 1. As well as accelerating the *in vitro* evolutionary process [10–12], the shuffling reaction is extremely flexible: many different pieces of genetic information may be included if they are available (see Figure 2; [13]). For example, Liu *et al.* [14] included degenerate oligonucleotides in their shuffling reaction in order to randomize amino acids believed, through structural studies, to be important for the substrate specificity of a tRNA synthase. Interestingly, only one of the five targeted residues was mutated in the enzyme showing highest activity against the new substrate.

Figure 1

In vitro recombination by DNA shuffling. Genes are fragmented and then reassembled by a reaction in which homologous fragments act as primers for each other.



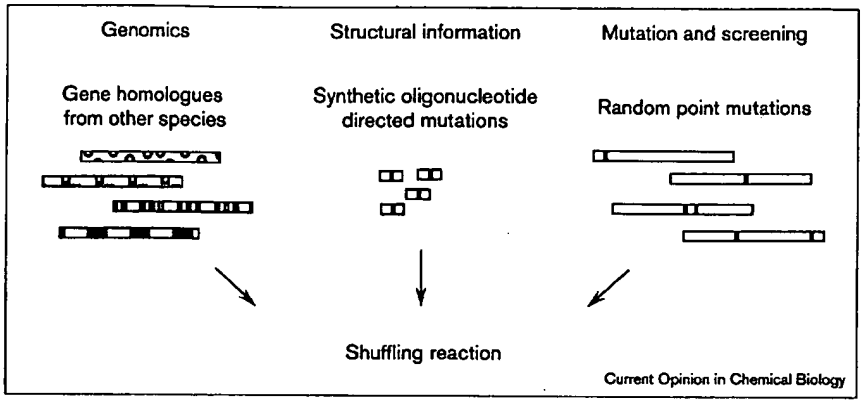
Many examples of successful directed evolution using DNA shuffling have been reviewed recently [15•,16•]. Last year, several additional formats were described for the *in vitro* [17,18] or *in vivo* [19] shuffling of genes. While these methods have not been thoroughly compared, they rely on the same underlying principle that the most efficient way to explore all of the possible combinations and permutations of sequences (i.e. sequence space) is by recombination of active variants.

Screening and selection

Natural evolution measures the fitness of variants by their ability to survive. In some cases, there are genetic selections that can be employed to make a cell's growth dependent on a particular improved function. Schellenberger's group [20•] recently selected for increased subtilisin production by making a target protein the sole source of nitrogen, performing the growth in hollow fibres to prevent cross-feeding. As an artificial selection system, phage display has been used to identify proteins that bind specific ligands. Catalytic proteins displayed on phage have also been selected, either by making infectivity dependent on formation of a covalent intermediate [21••], or by requiring enzyme activity to release the phage from a solid matrix [22•]. Both of these methods only require a single catalytic event, so are unsuitable for quantitative measurements.

Directed evolution has been used to enhance lipase enantioselectivity. Lipases accept a wide variety of non-natural esters, so lipases that are able to discriminate between stereoisomers allow the production of optically pure compounds useful in pharmaceutical and fine chemical manufacture. One group used a microtitre-based absorbance assay in which the esterase activity of lipase variants was measured against the *R* and *S* forms of *p*-nitrophenyl 2-methyldecanoate. Four cycles, testing 1,000 lipase mutants per cycle, increased the enantioselectivity from 2% enantiomeric excess (ee) to 81% ee in favor of the *S* configuration [23]. A second group evolved an

Figure 2



The shuffling reaction is extremely flexible. Positive variants resulting from random mutation and selection can be recombined with sequence information obtained computationally. Genomics allow the inclusion of related genes from other species and structural information can be used to design synthetic oligonucleotides for making specific changes or to randomize targeted regions of a protein.

enzyme to hydrolyze an ester for production of an intermediate in epithilone synthesis. The initial screen for this enzyme was performed by including both the enzyme substrate and a pH indicator in agar plates. Bacterial colonies expressing an enzyme able to hydrolyze the ester were identified by a change in the colour of the indicator, since acid is released when esters are hydrolyzed. Colonies selected by this screen were then picked and tested for their biotransformation activity and stereoselectivity by measuring the optical rotation of the products [24•]. While individual screens will always depend on the reactions being catalyzed, this strategy of tiered screening in which a primary, relatively inaccurate assay is used to select a small number of clones that are then subjected to more detailed analysis (see Figure 3) is an extremely powerful general technique.

It is also possible to perform an entire selection *in vitro*. As an example, a library of genes was transcribed and translated in compartments formed in a water/oil emulsion. Active DNA methyltransferase *Hae*III enzymes methylated the genes that encoded them, thereby protecting the DNA from subsequent *Hae*III digestion [25••]. By using such a system, cloning or transformation of the library is not required, so much larger libraries can be screened. Further advances such as coupled reactions leading to gene modification and sorting of intact compartments based on fluorescence would help make *in vitro* enzyme production and testing a very powerful methodology.

Using natural diversity

In addition to developing screening strategies that allow greater numbers of mutants to be screened, directed evolution can be optimized by building protein libraries that contain the maximum number of active (and different) members. Until this year, single genes were used as starting points for DNA shuffling and variants, arising by point mutation, were very similar in sequence to the parent gene. Another approach uses principles similar to those of the mammalian immune system. Antibodies capable of binding essentially any epitope with nanomolar association constants are generated by recombinination between a few thousand sequences, followed by 'affinity maturation' by point mutation [26]. Enzyme catalysis results from binding to and stabilizing the relevant transition-state analogue [27], so it should be possible to harness such a system to produce enzymes [28]. Antibodies have evolved as rigid binding molecules, however, and catalytic antibodies are selected solely by their abilities to bind transition-state analogues rather than other enzymatically essential functions such as substrate binding and product release. They are thus generally much less active as catalysts than proteins that have evolved as enzymes.

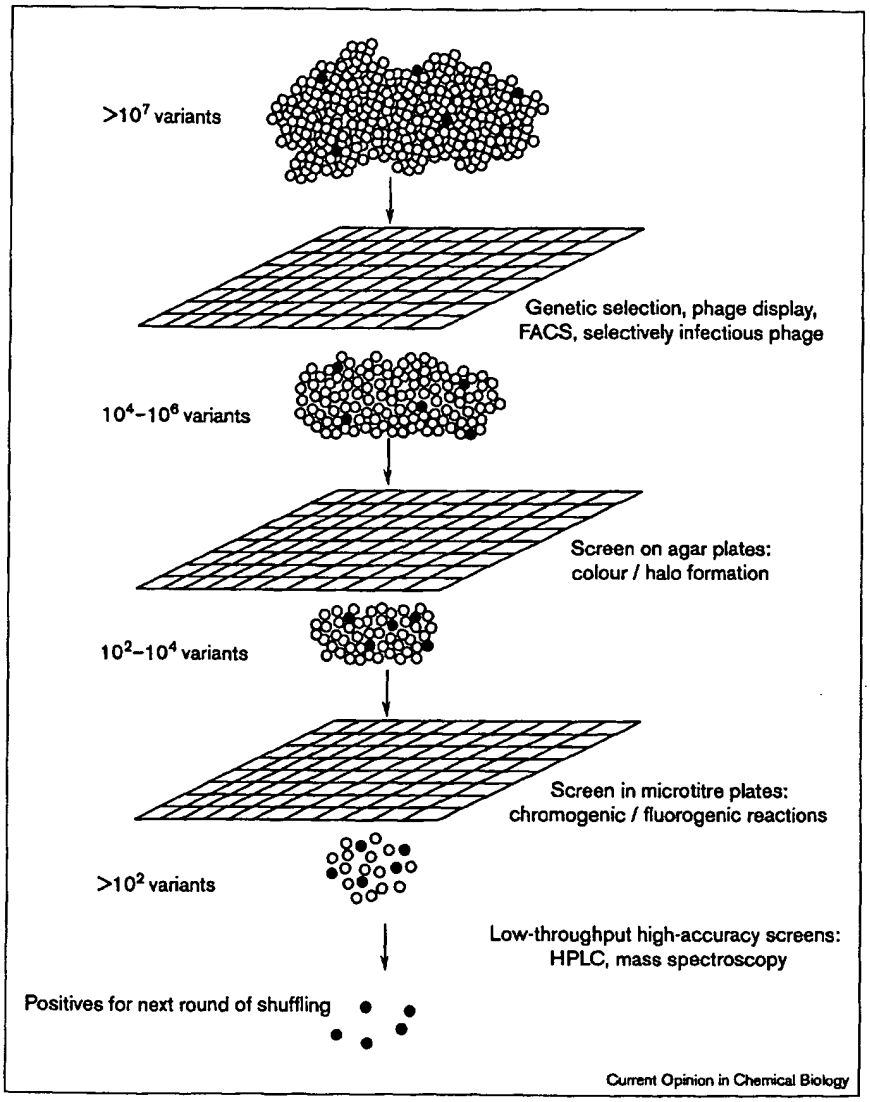
Instead of trying to turn antibodies into catalysts, DNA shuffling can be used to mimic the immune system's incredibly powerful diversity-generating process, by recombining genes with one another. In the first example of 'DNA family shuffling', four different β-lactamase genes were shuffled together to produce a chimera with 270-fold greater resistance to moxalactam than the best parental enzyme [29•]. The chimeric enzyme produced in this experiment differed from each parent by at least 100 amino acids (Figure 4), yet was still a fully functional cephalosporinase. Like antibody 'diversity' regions, sequences that occur in naturally existing enzymes have already been tested for their ability to function within the context of the protein's overall structure. Recombining natural blocks of sequence with each other allows a broad region of functional sequence space to be sampled sparsely.

Protein chimeras may differ dramatically from all their parents

Where an active site lies at the interface between folding subdomains, exchanging these subdomains will alter the shape of the active site. For example, swapping domains between coagulation factor X and trypsin produced a serine protease with broadened substrate selectivity [30•]. The activities of chimeric enzymes are often not predictable simply by comparing those of the parent enzymes,

Figure 3

Tiered screening. Variants are tested by a series of assays that are successively more accurate and more time- and labour-intensive. It is important to ensure that the higher capacity assays correlate well with the desired final activity. FACS, fluorescence-activated cell sorting; HPLC, high-pressure liquid chromatography.

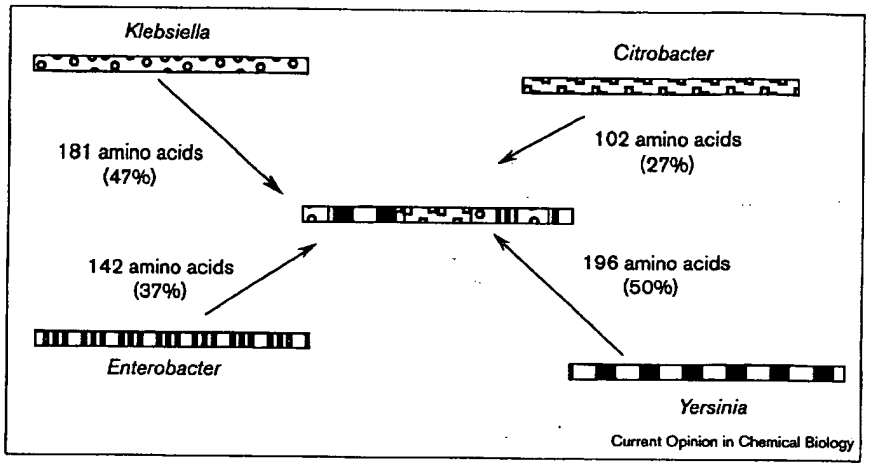


as was found for chimeras between two human blood group glycosyl transferases that were shown to be functionally interconvertible by changing only four amino acids. Parental enzyme A transfers *N*-acetylgalactosamine to a disaccharide acceptor, whereas enzyme B transfers galactose. Replacement of Arg176 in enzyme A with the Gly176 of enzyme B resulted not in increases in B-like activities, but in a fourfold higher k_{cat}/K_M for the enzyme A substrate (i.e. *N*-acetylgalactosamine) [31••].

Altered substrate specificities have also been produced by random recombination of sequences followed by screening. Biphenyl dioxygenases initiate the degradation of polychlorinated biphenyls, and their congener substrate specificities are determined by the large terminal subunit [32]. DNA shuffling of two such dioxygenases produced chimeras with a different substrate range from either parent, enhanced degradation of biphenyl compounds and even novel oxygenation activity for single aromatic hydrocarbons [33,34••].

Random chimeras have also been made *in vivo* between two staphylococcal lipases with differing chain-length selectivities and phospholipase activities. Novel enzymes were found that possessed both combinations of and absolute levels of these activities that differed from both parents in ways that were often surprising [35•]. For example, one chimera in which a block comprising 20% of the enzyme with no chain-length selectivity was incorporated into the enzyme with a strong preference for short-chain fatty acids unexpectedly resulted in an enzyme with twofold increased activity (relative to the best parent) against the long-chain ester *p*-nitrophenyl palmitate.

Figure 4



Mutational distances of chimeric β-lactamase with 270-fold improved moxolactamase activity from its four parents. Distances from each parent are given in number of amino acids, and in the percentage of residues that this represents. The chimera differs by 102 amino acids, that is 27% of positions, from its closest parent (the *Citrobacter* enzyme). It would not be possible to make 102 random changes without inactivating the enzyme. Thus recombination of natural diversity allows functional sequence space to be sampled much more broadly and sparsely than sequential point mutations from a single starting sequence.

Recursive cycles of shuffling using multiple parents has been performed by Christians *et al* [36••]. By recombining two Herpes Simplex Virus thymidine kinase genes and robotically screening for variants that were better able to phosphorylate the therapeutic nucleotide analogue AZT, the concentration of AZT required to inhibit cell growth was reduced 32-fold relative to that required with the best parent. The resulting enzyme was a chimera that had undergone ten cross-over events between the two parental genes, and had also accumulated five point mutations, leading to a protein differing by 22 amino acids from the closest parent. The process of recombination between different but functional parents to make large changes in sequence, coupled with point mutagenesis to fine-tune the activity of the protein, is highly analogous to the process of antibody generation and maturation.

Directed searches for novel protein activities

Although it is possible to modify the physical properties of an enzyme, such as thermostability or activity in organic solvent, by screening for sequential improvements in these properties [37–39], modification of one property by single point mutations can often compromise another desired characteristic [40•]. From the results discussed above, we would predict that by recombining sequences found in nature, it should be possible to discover enzymes possessing all combinations of properties of the individual parents, as well as improvements over any of the parents.

The classification of enzymes into superfamilies that appear to be related by a common chemical strategy for stabilizing the transition state for the formation of a reactive intermediate suggests a mechanism by which nature may evolve novel catalytic functions [41]. Is it possible to make such changes in the laboratory? It may not be possible to make a graded change from one reaction to another. By making structural comparisons between an oleate desaturase and an oleate hydroxylase, Broun *et al.* [42••] have shown that four amino acid changes in the desaturase can convert it to a hydroxylase and changing six residues in the hydroxylase result in desaturase activity. Making these changes by sequential point mutagenesis would not be possible because the single or double mutants do not possess intermediate activities. The exchange of blocks of amino acids made possible by family shuffling, however, offers a possible route to completely novel substrate specificities. Enzyme libraries constructed from relatively small families of homologous genes are likely to contain not only a range of substrate specificities, but also a variety of physical properties and even new catalytic activities. These libraries can then serve as sources of diversity themselves, providing the starting points for further directed evolution in many different directions.

Conclusions

By copying the natural mechanisms by which even existing diversity can be recombined, DNA shuffling can be used to generate high-quality libraries of novel proteins. Chimeras between naturally occurring enzymes that differ by only a few amino acids often possess activities that are significantly different from their parents. By screening these libraries using innovative high-throughput assay techniques, it is possible to identify enzymes with new catalytic functions and physical properties.

Acknowledgements

We thank SunAi Raillard, Claus Krebber, Andreas Crameri, Abby Dernburg and Matt Tobin for stimulating discussions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Robertson DE, Mathur EJ, Swanson RV, Marrs BL, Short JM: **The discovery of new biocatalysts from microbial diversity.** *Soc Indust Microbiol News* 1996, 46:3-8.

2. Madigan MT, Marrs BL: **Extremophiles.** *Sci Am* 1997, **276**:82-87.

3. Rubingh DN: **Protein engineering from a bioindustrial point of view.** *Curr Opin Biotechnol* 1997, **8**:417-422.

4. Hellinga HW: **Rational protein design: combining theory and experiment.** *Proc Natl Acad Sci USA* 1997, **94**:10015-10017.

5. Shao Z, Arnold FH: **Engineering new functions and altering existing functions.** *Curr Opin Struct Biol* 1996, **6**:513-518.

6. Kelly K: *Out of Control: The Rise Of Neo-Biological Civilization.* Menlo Park, California: Addison–Wesley; 1994.

7. Holland JH: *Adaptation In Natural And Artificial Systems.* Ann Arbor, MI: University of Michigan Press; 1975.

8. Forrest S: **Genetic algorithms: principles of natural selection applied to computation.** *Science* 1993, **261**:872-878.

9. Gibson JM: **Simulated evolution and artificial selection.** *Biosystems* 1989, **23**:219-229.

10. Stemmer WP: **Rapid evolution of a protein *in vitro* by DNA shuffling.** *Nature* 1994, **370**:389-391.

11. Stemmer WPC: **Searching sequence space.** *Biotechnology* 1995, **13**:549-553.

12. Moore JC, Jin HM, Kuchner O, Arnold FH: **Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences.** *J Mol Biol* 1997, **272**:336-347.

13. Stemmer WP: **DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution.** *Proc Natl Acad Sci USA* 1994, **91**:10747-10751.

14. Liu DR, Magliery TJ, Pastmak M, Schultz PG: **Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.** *Proc Natl Acad Sci USA* 1997, **94**:10092-10097.

15. • Giver L, Arnold FH: **Combinatorial protein design by *in vitro* recombination.** *Curr Opin Chem Biol* 1998, **2**:335-338.
This is a good summary of several examples of protein evolution by recombination-based methods.

16. • Harayama S: **Artificial evolution by DNA shuffling.** *Trends Biotechnol* 1998, **16**:76-82.
This is a review of DNA shuffling examples.

17. Shao Z, Zhao H, Giver L, Arnold FH: **Random-priming *In vitro* recombination: an effective tool for directed evolution.** *Nucleic Acids Res* 1998, **26**:681-683.

18. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH: **Molecular evolution by staggered extension process (StEP) *in vitro* recombination.** *Nat Biotechnol* 1998, **16**:258-261.

19. Zhang Y, Buchholz F, Muyrers JPP, Stewart AF: **A new logic for DNA engineering using recombination in *Escherichia coli*.** *Nat Genet* 1998, **20**:123-128.

20. • Naki D, Paech C, Granshaw G, Schellenberger V: **Selection of a subtilisin-hyperproducing *Bacillus* in a highly structured environment.** *Appl Microbiol Biotechnol* 1998, **49**:290-294.
This is a selection designed to link increased protease activity to faster growth rate. The authors identified a mutant with increased expression, but is also a more active protease.

21. •• Gao C, Lin CH, Lo CHL, Mao S, Wirsching P, Lerner RA, Janda KD: **Making chemistry selectable by linking it to infectivity.** *Proc Natl Acad Sci USA* 1997, **94**:11777-11782.
Linking genotype to phenotype is essential for evolution. Here, phage infectivity requires a catalytically active protein to be fused to the gene III protein. The authors were only able to detect single turnover events, but the technique represents a potential way to screen large libraries for novel catalytic function.

22. • Pedersen H, Holder S, Sutherlin DP, Schwitter U, King DS, Schultz PG: **A method for directed evolution and functional cloning of enzymes.** *Proc Natl Acad Sci USA* 1998, **95**:10523-10528.
The authors describe another way of using phage display to identify catalysts. They showed that binding to or release from a physical support is dependent on a single turnover event.

23. Reetz MT, Zonta A, Schimossek K, Liebeton K, Jaeger K-E: **Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution.** *Agnew Chem Int Ed Engl* 1997, **36**:2830-2832.

24. • Bornscheuer UT, Altenbuchner J, Meyer HH: **Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epithilones.** *Biotechnol Bioeng* 1998, **58**:554-559.
A tiered screen is described, and directed evolution of enantioselectivity is presented.

25. •• Tawfik DS, Griffiths AD: **Man-made cell-like compartments for molecular evolution.** *Nat Biotechnol* 1998, **16**:652-656.
An elegant *in vitro* transcription, translation and selection system is described in which cloning and transformation are things of the past. Some major limitations still need to be overcome to allow general applicability, but one day all enzymes will be evolved like this.

26. Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302**:575-581.

27. Pauling L: **Molecular architecture and biological reactions.** *Chem Eng News* 1946, **24**:1375-1377.

28. Jencks WP: *Catalysis In Chemistry And Enzymology.* New York: McGraw Hill; 1969.

29. • Crameri A, Raillard S-A, Bermudez E, Stemmer WPC: **DNA shuffling of a family of genes from diverse species accelerates directed evolution.** *Nature* 1998, **391**:288-291.
Here, generation of an improved variant enzyme that is very different from it's parents is described. The authors demonstrate that chimeragenesis is a good way to sample broad areas of sequence space sparsely.

30. • Hopfner K-P, Kopetzki E, Kreße G-B, Bode W, Huber W, Engh RA: **New enzyme lineages by subdomain shuffling.** *Proc Natl Acad Sci USA* 1998, **95**:9813A.
A deliberate attempt is made to change the shape of the catalytic cleft by making a chimera with folding subdomains from different proteases. Novel but unpredicatable properties result.

31. •• Seto NOL, Palcic MM, Compston CA, Li H, Bundle DR, Narang SA: **Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes.** *J Biol Chem* 1997, **272**:14133-14138.
Exchange of very limited natural diversity between two functionally distinct enzymes produces chimeras with unexpected activities is described.

32. Erickson BD, Mondello FJ: **Enhanced biodegradation of polychlorinated biphenyls after site-directed mutagenesis of a biphenyl dioxygenase gene.** *Appl Environ Microbiol* 1993, **59**:3858-3862.

33. Kimura N, Nishi A, Goto M, Furukawa K: **Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally.** *J Bacteriol* 1997, **179**:3936-3943.

34. •• Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K: **Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase.** *Nat Biotechnol* 1998, **16**:663-666.
DNA shuffling of two highly homologous large subunits of multisubunit enzymes is described. Chimeras possess novel substrate preferences and are still able to interact with the other components of the enzyme system.

35. • van Kampen MD, Dekker N, Egmond MR, Verheij HM: **Substrate specificity of *Staphylococcus* hyicus lipase and *Staphylococcus aureus* lipase as studied by *in vivo* chimeragenesis.** *Biochemistry* 1998, **37**:3459-3466.
The authors describe the use of chimeras to crudely correlate primary sequence with enzyme substrate specificity. The main conclusion drawn is that chimeric proteins can exceed the sum of their parts. Quite right!

36. •• Christians FC, Scapozza L, Crameri A, Folkers G, Stemmer WPC: **Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling.** *Nat Biotechnol* 1999, **17**:259-264.
Chimeragenesis, multiple rounds of DNA shuffling and a robotic screen for improved function are described. Also, a nice test of how many mutants one should take from one round to the next is discussed.

37. Zhao H, Arnold FH: **Directed evolution converts subtilisin E into a functional equivalent of thermitase.** *Protein Eng* 1999, **12**:47-53.

38. Akanuma S, Yamagishi A, Tanaka N, Oshima T: **Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from Bacillus subtilis by experimental evolution.** *Protein Sci* 1998, **7**:698-705.

39. Buchholz F, Angrand P-O, Stewart AF: **Improved proerties of FLP recombinase evolved by cycling mutagenesis.** *Nat Biotechnol* **1998, 16:657-662.**

40. • Patkar S, Vind J, Kelstrup E, Christensen MW, Svendsen A, Borch K, Kirk O: **Effect of mutations in *Candida antarctica* B lipase.** *Chem Phys Lipids* **1998, 93:95-101.**
The authors juggle with thermostability, enantioselectivity, specific activity and oxidation stability of a lipase. They conclude that changes that improve one property degrade another.

41. Gerlt JA, Babbitt PC: **Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis.** *Curr Opin Chem Biol* **1998, 2:607-612.**

42. •• Broun P, Shanklin J, Whittle E, Somerville C: **Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids.** *Science* **1998, 282:1315-1317.**
Small numbers of amino acid substitutions change not only the substrate specificity of an enzyme, but also the reaction catalysed. Watch in awe as a desaturase becomes a hydroxylase.

APPENDIX E

Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling

Fred C. Christians[1,3], Leonardo Scapozza[2], Andreas Crameri[1], Gerd Folkers[2], and Willem P.C. Stemmer[1]*

*[1]Maxygen, Inc., 3410 Central Expressway, Santa Clara, CA 95051. [2]Department of Pharmacy, Swiss Federal Institute of Technology (ETH), CH-8057, Zürich, Switzerland. [3]Current address: Affymetrix, Inc., 3380 Central Expressway, Santa Clara, CA 95051. *Corresponding author (e-mail: pim_stemmer@maxygen.com).*

Received 24 August 1998; accepted 15 January 1999

The thymidine kinase (TK) genes from herpes simplex virus (HSV) types 1 and 2 were recombined in vitro with a technique called DNA family shuffling. A high-throughput robotic screen identified chimeras with an enhanced ability to phosphorylate zidovudine (AZT). Improved clones were combined, reshuffled, and screened on increasingly lower concentrations of AZT. After four rounds of shuffling and screening, two clones were isolated that sensitize *Escherichia coli* to 32-fold less AZT compared with HSV-1 TK and 16,000-fold less than HSV-2 TK. Both clones are hybrids derived from several crossover events between the two parental genes and carry several additional amino acid substitutions not found in either parent, including active site mutations. Kinetic measurements show that the chimeric enzymes had acquired reduced K_M for AZT as well as decreased specificity for thymidine. In agreement with the kinetic data, molecular modeling suggests that the active sites of both evolved enzymes better accommodate the azido group of AZT at the expense of thymidine. Despite the overall similarity of the two chimeric enzymes, each contains key contributions from different parents in positions influencing substrate affinity. Such mutants could be useful for anti-HIV gene therapy, and similar directed-evolution approaches could improve other enzyme–prodrug combinations.

Keywords: random mutagenesis, sexual PCR/DNA shuffling, suicide gene, gene transfer, zidovudine

DNA shuffling combined with an activity screen can accelerate directed evolution of desired traits. In DNA shuffling of single starting sequences[1,2], diversity is introduced as a library of random point mutations. Following screening, improved clones are reshuffled to recombine useful mutations in additive or synergistic ways, in effect mimicking the process of natural sexual recombination. This method has been used to enhance enzyme activity significantly[2], change substrate specificity[3], and improve protein folding[4,5]. A recent adaptation called family shuffling[6] allows two or more naturally occurring sequences to be used as the starting genetic material. In this method, homologous genes are mixed, randomly fragmented, and recombined using conditions that permit annealing and extension of nonidentical complementary strands. Such related genes provide functional diversity as opposed to simply random mutations, an important distinction because only about 1% of random mutations in a gene are beneficial. This use of functional diversity, in the form of related genes already subjected to millions of years of natural selection, bypasses the limitations of natural species barriers and allows rapid searching of large and diverse regions of sequence space. DNA family shuffling has been used to evolve β-lactamase from four starting genes[6] and biphenyl dioxygenase from two genes[7].

The lower substrate specificity of herpesviral thymidine kinases (TKs) relative to the human TK provides the basis for the efficacy of nucleoside analogs such as acyclovir and ganciclovir, both as antiviral agents[8] and in suicide gene therapy[9]. The most active TK is that of herpes simplex virus type 1 (HSV-1), which is being evaluated in numerous clinical trials. The use of HSV-1 TK in bolstering the effectiveness of zidovudine (AZT) in anti-HIV therapy has also been considered[10,11]. We sought to improve the ability of herpesvirus TK to phosphorylate AZT. The TK genes from HSV-1 and HSV-2 are the two most closely related herpesvirus TK genes (78% DNA identity) in this highly divergent family[12]. We used the DNA family shuffling method to create a library of HSV-1/HSV-2 TK chimeras, which were tested in a high-throughput robotic screen to identify clones capable of sensitizing *Escherichia coli* to AZT. Clones with improved function were reshuffled and rescreened for a total of four cycles to evolve chimeras with greatly enhanced activity on AZT.

Results

Family shuffling and screening. Seventeen of 20 randomly picked clones from the initial shuffled pool showed novel restriction patterns when the TK genes were treated with the restriction endonuclease *Sau*3AI, an indication of efficient recombination between the two parental genes. About 20% of the clones in this initial library retained sufficient TK activity to permit complementation of *E. coli* KY895, a TK-deficient mutant. About 10,000 clones per cycle were screened using a robotic colony picker to identify clones that conferred enhanced AZT sensitivity to KY895. Library transformants were plated in replicate on Luria-Bertani media with or without AZT. Clones that were selectively growth inhibited by AZT were chosen for the next cycle, while clones that grew poorly on LB alone were not chosen. At no time were the clones placed under selective pressure for TK activity on thymidine, unlike previous directed evolution of TK[13–15]. We used the TK-defective strain for selection to eliminate background TK activity, so that activation of the prodrug AZT could be attributed to the introduced TK genes; KY895 itself showed no sensitivity to AZT.

Table 1. Screening shuffled TK libraries for AZT sensitivity.

Round	Number of clones screened	Number sensitive	AZT (ng/ml)
HSV-1 TK	–	–	500
HSV-2 TK	–	–	4,000
1	11,520	124	50
2	10,752	47	10 (= tier 2)
		6	5 (= tier 1)
3[a]	11,520 (tier 1)	3	5
		4	2.5
	11,520 (tiers 1+2)	21	5
		28	2.5
		1	1.5
4	8,991	1	1

[a]In round 3, two separate shuffling reactions were done: one with the tier 1 clones from round 2 (sensitive to 5 ng/ml AZT), and one reaction with a mixture of all tier 1 and 2 clones.

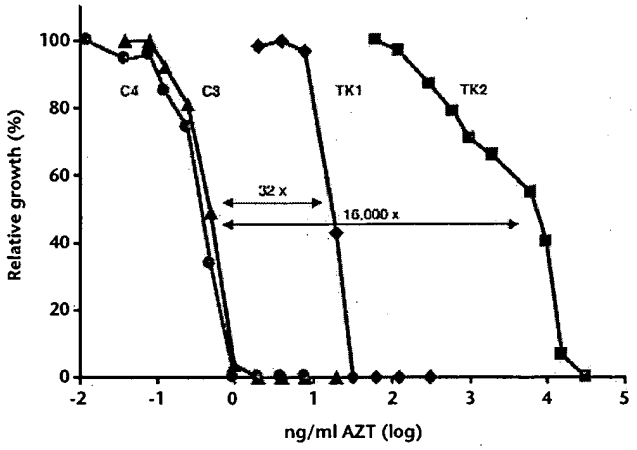


Figure 1. Titration of growth inhibition of TK-expressing KY895 cells by AZT. Dilute cultures were plated on LB media containing various concentrations of AZT and incubated overnight. Growth was determined by measuring the optical density of the bacteria after resuspension in saline and is expressed relative to cultures grown on AZT-free plates. Abbreviations: C4 = cycle 4 TK; C3 = cycle 3 TK; TK1 = HSV-1 TK; TK2 = HSV-2 TK.

Results of the screening assay are shown in Table 1. The initial parental clones in this assay failed to grow at AZT concentrations above 500 ng/ml (HSV-1 TK) or 4000 ng/ml (HSV-2 TK). One hundred twenty-four clones from the first round of shuffling performed at least 10-fold better than HSV-1 TK. Restriction fragment length polymorphism (RFLP) analysis of 20 of these improved clones revealed a minimum of 12 different patterns, a finding suggesting a variety of solutions, including those incorporating genetic contributions from HSV-2 TK. This result indicated that the diversity provided by a less active variant (HSV-2 TK) can be used to improve a more active variant (HSV-1 TK) by molecular breeding. Shuffling together the 124 improved clones from the first round resulted in further improvement, with six clones in the second round being sensitive to 5 ng/ml AZT (tier 1) and an additional 47 clones sensitive to 10 ng/ml (tier 2). RFLP analysis showed that these sequences still comprised a family of diverse chimeras.

In the third cycle of molecular breeding, we tested whether additional diversity can yield improvements in fitness, even when provided by less active sequences. Two otherwise identical shuffling reactions were compared, one involving only the best six clones from cycle 2 (tier 1), and the other involving all of the 53 best cycle 2 clones (tiers 1+2). Equal numbers of clones from each set were screened for AZT sensitivity. The findings (Table 1) clearly show that the additional diversity from tier 2 enhanced the shuffling result: Only four clones from the tier 1 reaction performed better than any of the parents, whereas 29 clones from the tier 1+2 reaction showed improvement. This included one clone, "cycle 3 TK," that was superior to any from the tier 1 reaction. The results of this experiment differ from those of a different, single gene shuffling system[16], in which it was found that the additional diversity provided by less highly evolved clones slowed the evolutionary process. We expect that the ideal parental pool size for recombination will vary in different systems.

For the fourth cycle we used a mixture of the 33 best clones from cycle 3, doped with a 10-fold smaller amount of DNA from the next-best clones. Primary screening of 8,991 clones revealed that the best clone, "cycle 4 TK," sensitized KY895 to 1 ng/ml AZT, a 500-fold improvement over HSV-1 TK and a 4,000-fold improvement over HSV-2 TK. In a more precise secondary screen based on a growth inhibition titration assay, the 50% inhibitory concentration of AZT for both the cycle 3 TK and cycle 4 TK clones was reduced 32-fold relative to HSV-1 TK and 16,000-fold relative to HSV-2 TK (Fig. 1).

Kinetics. The HSV-1, HSV-2, cycle 3 TK, and cycle 4 TK proteins were purified and analyzed for their ability to phosphorylate thymidine and AZT. All four proteins were expressed at equal levels. The calculated kinetic parameters of the cycle 3 TK and cycle 4 TK were similar (Table 2). One key distinction of the evolved enzymes was an approximately threefold or 10-fold reduction in K_M for AZT relative to HSV-1 TK and HSV-2 TK, respectively. The affinity of the four enzymes for AZT roughly correlated with their relative abilities to sensitize *E. coli* to AZT (Fig. 1). Another kinetic difference was the shift of specificity from thymidine to AZT (the k_{cat}/K_M of thymidine relative to that of AZT). The evolved enzymes showed only a 1.6-fold overall preference for thymidine, a reduction of 44-fold compared with HSV-1 TK and sevenfold compared with HSV-2 TK. The in vitro kinetics indicated that the gain in the ability to catalyze the phosphorylation of AZT came at the expense of activity on thymidine. There was no absolute correlation between this specificity factor measured in vitro and the ability to sensitize *E. coli* to AZT (Fig. 1), because HSV-2 TK showed very low AZT sensitization despite intermediate specificity.

Table 2. Kinetics of thymidine and AZT phosphorylation by purified TK proteins.

	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}\mu M^{-1}$)	Specificity = $\frac{k_{cat}/K_M \text{ (thymidine)}}{k_{cat}/K_M \text{ (AZT)}}$
Thymidine				
HSV-1 TK	0.7 ± 0.36	3.6 ± 0.26	5.1	71
HSV-2 TK	4.8 ± 0.43	0.55 ± 0.02	0.11	11
Cycle 3 TK	9.3 ±3.6	1.5 ± 0.20	0.16	1.6
Cycle 4 TK	3.5 ± 2.4	0.59 ± 0.10	0.17	1.6
AZT				
HSV-1 TK	16 ± 1.4	1.2 ± 0.04	0.073	
HSV-2 TK	49 ± 2.2	0.51 ± 0.01	0.010	
Cycle 3 TK	4.9 ± 1.9	0.50 ± 0.15	0.10	
Cycle 4 TK	5.9 ± 2.3	0.63 ± 0.06	0.11	

Amino acid sequence. The deduced sequences of cycle 3 TK and cycle 4 TK indicate that chimeras were generated, consisting of sequence from the HSV-1 and HSV-2 TKs as well as additional point mutations (Fig. 2). The sequence of cycle 3 TK indicates that at least six DNA crossover events occurred during the DNA shuffling. The six amino acids that cannot be attributed to either parent were a consequence of the low-level point mutagenesis provided by the chosen shuffling conditions. The deduced protein sequence of cycle 3 TK

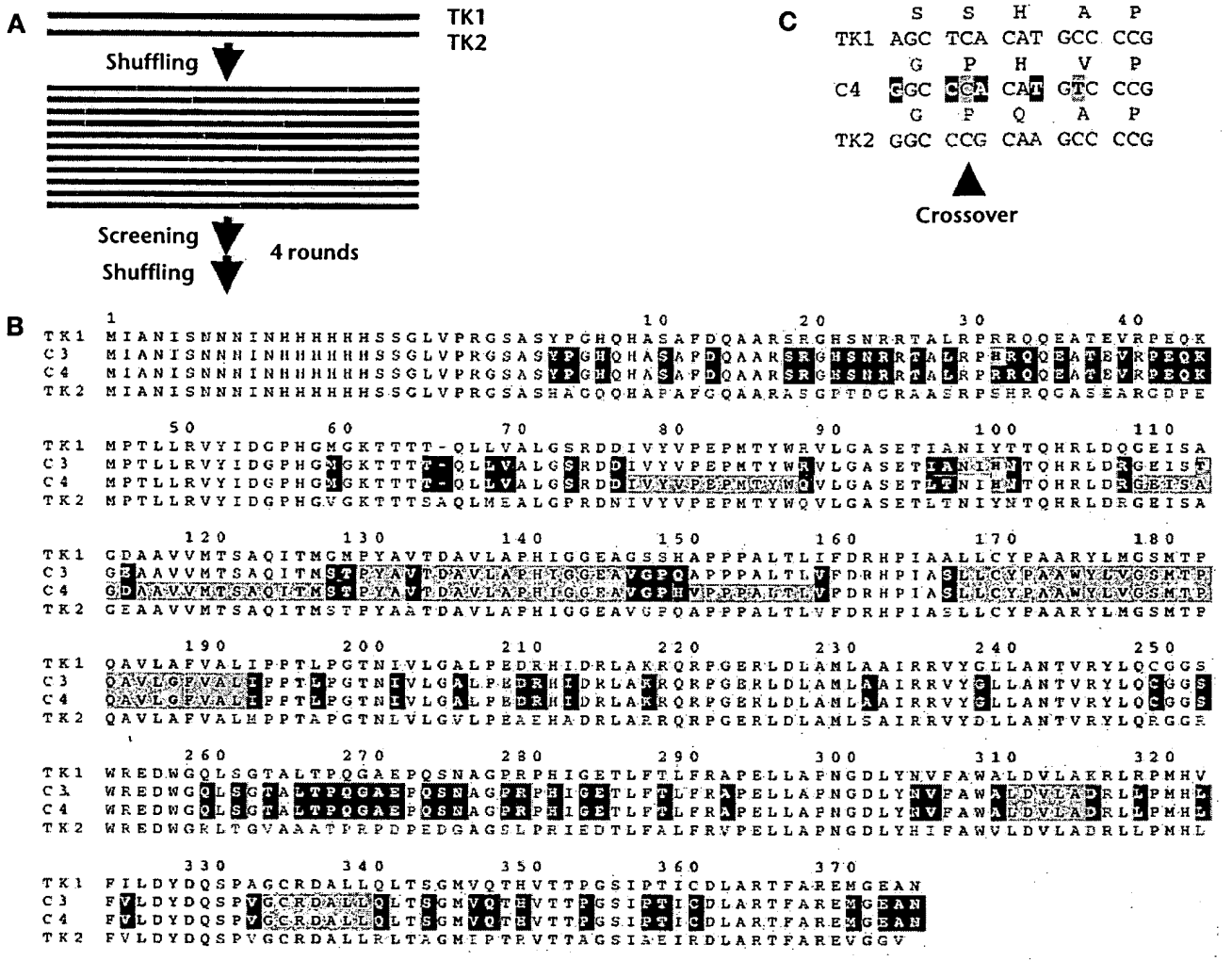


Figure 2. (A) Shuffling of the two parental TK genes creates a library of chimeric genes. Clone C3 was obtained after three cycles of screening and reshuffling, and clone C4 after four cycles. Abbreviations are the same as for Figure 1. (B) Deduced amino acid sequence of the two evolved clones and the parents. The shaded residues indicate homology to one parent or the other (green = TK1, red = TK2), point mutations (yellow), and regions of homology to both parents (gray) in which the DNA crossovers occurred to create the chimeras. The numbers above the sequences are those of HSV-1 TK, and disregard the 25-amino acid N-terminal tag (see Experimental protocol). (C) The DNA sequences corresponding to amino acid residues 149–153 suggest a crossover event in C4 with only 1 bp of exact homology between nonidentical sequences.

shows that it is 94% identical to HSV-1 TK (22 differences) and 77% identical to HSV-2 TK (86 differences). The sequence of cycle 4 TK is quite similar to that of cycle 3 TK (98% identical) but with some different crossovers and some different point mutations. At least 10 DNA crossover events occurred during the four rounds of shuffling to create cycle 4 TK. One crossover could be pinpointed to 1 bp (Fig. 2C) and another to 3 bp, demonstrating that fine-grained recombination is achievable by DNA shuffling. Such crossovers are made possible by flanking homology. Three of the point mutations (Y101H, R176W, and M179V) found in both enzymes are in the active site (see below).

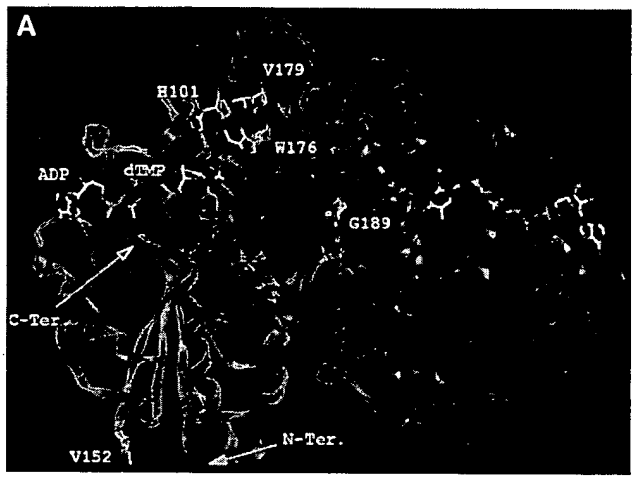
Discussion

To understand better how the evolved mutations improve the enzyme activity of TK, three-dimensional structures were modeled for cycle 4 TK and HSV-2 TK bound to dTMP and AZT, and for HSV-1 TK bound to AZT. These models were based on the crystal structure coordinates of HSV-1 TK[17]. The location of the nonconserved amino acids and the distribution of segments identical to HSV-1 TK and HSV-2 TK are shown in Figure 3A. Figure 3B shows a comparison of the binding sites of cycle 4 TK and HSV-1 TK. Figure 4 shows a comparison of the predicted hydrogen bond patterns of thymidine and AZT bound to HSV-1 TK, HSV-2 TK, and cycle 4 TK. At this time, conclusions are speculative because of amino acid differences among the modeled enzymes, including substitutions far from the active site, as well as the relatively low resolution of the original crystal structure (2.75 Å). However, the models yield several interesting predictions supported by the kinetic and *E. coli* sensitization data and also shed light on the molecular events that occurred in the DNA family shuffling reactions.

One of these predictions is that the two mutations, Y101H and R176W, in cycle 4 TK play prominent roles in the reorganization of the active site. These two positions participate in forming the thymidine-binding site[17,18], the predicted shape and volume of which differ significantly between cycle 4 TK and HSV-1 TK (Fig. 3B). The calculated GRID energy maps suggest that the mutation Y101H extends the binding site around the 3′ position of thymidine, an extension that accommodates the bigger azido substituent of AZT. At the same time, the R176W mutation decreases the available space around position N3 of dT, thereby possibly allowing a tighter nonspecific

binding of the ligand. The mutation of Met179 into the less bulky valine compensates for the active-site sterical hindrance created by the Arg176 mutation. In addition, the charged amino acid arginine, which usually mediates an altered pH-dependency of the mutant enzyme due to a pK shift of the nearby His101, can be altered by substitution of the arginine with the more hydrophobic trytophan. This prediction was supported by the in vitro kinetics and pH dependency of the Try101→His single mutant engineered from HSV-1 TK (data not shown).

The mutations Try101→His and Arg176→Trp also induced the reorganization of the water-mediated H-bond network found in the crystal structure (Fig. 4) and allowed us to speculate on the relation-



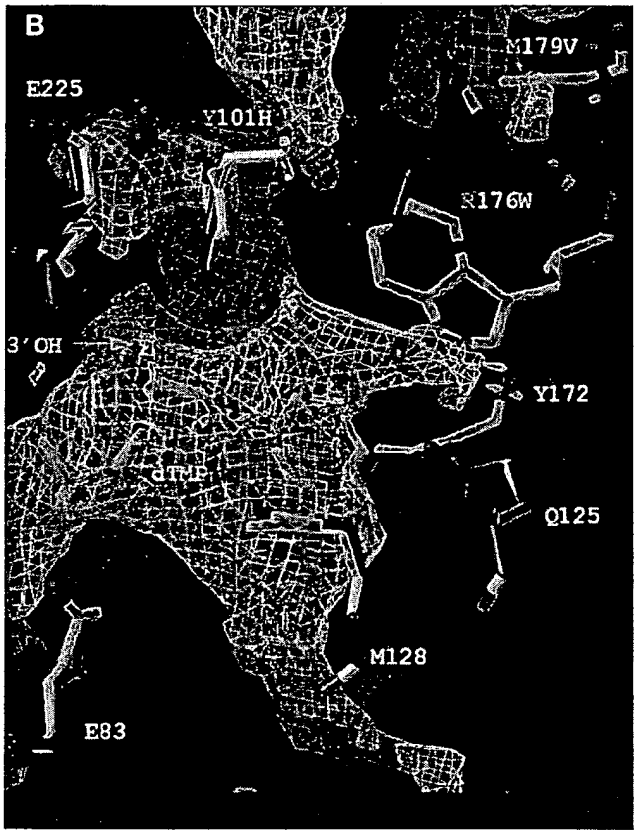


Figure 3. (A) Minimized three-dimensional model of cycle 4 TK with bound ADP and dTMP showing the location of the five nonconserved amino acids (capped sticks). A water molecule within the binding site responsible for water-mediated H-bonds between dTMP and the protein is illustrated as a red sphere. The backbone of the protein is indicated as a color-coded shaded ribbon. In green are shown segments in common between cycle 4 TK and HSV-1 TK. The red segments denote identity between HSV-2 TK and cycle 4 TK. The amino acid point mutations of cycle 4 TK are displayed in yellow. The backbone of the C2-symmetry–related monomer is shown as a magenta-shaded ribbon allowing visualization of the interface of the dimer. (B) GRID energy contour maps for HSV-1 TK (orange) and cycle 4 TK (green). The maps illustrate the interaction energy between a probe with the characteristics of an sp N with lone pair atom and TK and are contoured at a level of -0.3 kcal/mol. dTMP and the surrounding amino acids are labeled and displayed as capped sticks. The position of the 3′-OH group of dTMP is marked. The atoms of cycle 4 TK are color-coded (C: white, N: blue, O: red, P: orange, S: yellow), whereas the carbon atoms of HSV-1 TK structure are in purple. Water molecules that are within the active site, forming water-mediated H-bonds with dTMP, or in the cavity nearby are shown as spheres (cyan for cycle 4 TK, red for HSV-1 TK).

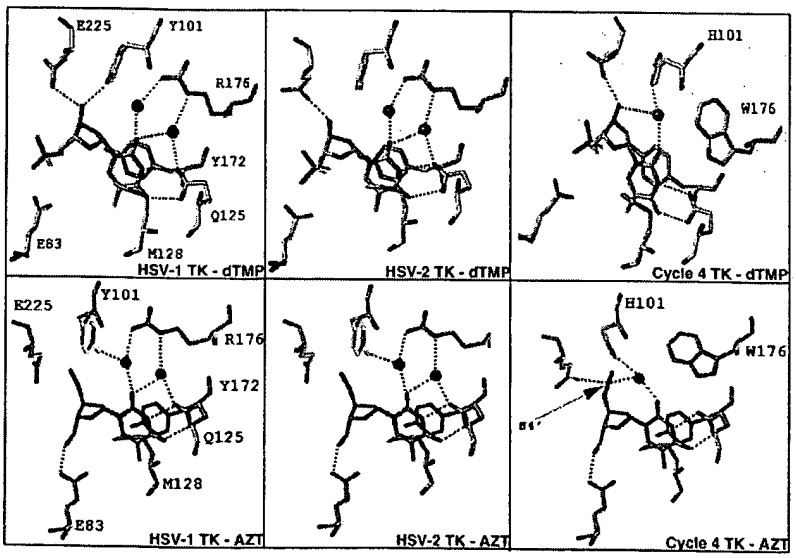


Figure 4. Comparison of the hydrogen-bonding patterns of the binding sites of HSV-1 TK, HSV-2 TK, and cycle 4 TK with dTMP and AZT. The substrates and amino acids are shown as capped sticks, water molecules are shown as spheres, and the hydrogen bond network is shown as dashed lines.

ship between the kinetic findings (Table 2) and structure. The lone water molecule remaining in the binding site of cycle 4 TK assumed a crucial role in binding the substrates. A hydrogen bond between Try101 and the 3′-OH group of thymidine-5′-monophosphate (dTMP) is present in HSV-1 TK and absent in HSV-2 TK because of the insertion of an alanine at position 66 in HSV-2 TK. The loss of this hydrogen bond, which is in close contact to the phosphate-binding region of the ATP binding site, corresponds to the K_M difference of one order of magnitude seen between HSV-1 TK and HSV-2 TK. The model of cycle 4 TK–dTMP predicts that the base ring and the 3′-OH of dTMP are still well fixed to the protein via hydrogen bonds. Two water-mediated H-bonds are lost relative to HSV-1 TK, which corresponds to the loss of one order of magnitude of binding affinity.

In addition, the model of HSV-1 TK–AZT shows that Glu225 and Try101 are pointing away from the azido group because of the limited space. Compared with the known crystal structure of HSV-1 TK–dTMP, two H-bonds are lost and no additional favorable interactions between the azido group and the protein are formed, in agreement with the measured increase of k_m and decrease of k_{cat} for AZT compared with dT. HSV-2 TK–AZT shows that Glu225 and Try101 do not interact with the azido group. Compared with HSV-2 TK–dTMP one hydrogen bond is lost, corresponding to the observed decreased binding affinity. In contrast, cycle 4 TK–AZT shows that the partially negatively charged oxygen of the water molecule and of the carboxyl group of Glu225 are interacting favorably with the positively charged N4′ of the azido group of AZT, while the H-bonds with the base are conserved. Thus, all interactions shown for thymidine are also present for AZT, explaining the unchanged binding affinity for AZT compared with thymidine, but also allowing cycle 4 TK to be less specific for thymidine and more specific for AZT compared with HSV-1 and HSV-2 TK.

The predictions of such rearrangements of hydrogen bond networks due to active-site changes, and their influences in the binding affinity of substrates are in agreement with the recently published structures of HSV-1 TK complexed with several uracil and guanine analogs[19]. In that report, the structures indicate that the general binding mode is conserved, that the differences in binding affinities are due to a variation of the hydrogen-bonding pattern, and that the change of conformation of one residue accommodates a larger substituent.

Cycle 3 TK was characterized at the same level as cycle 4 TK and was found to be substantially different despite an overall homology of 98%. Of the eight amino acid differences in the two enzymes, our models suggest that those at positions 89, 97, and 98 are the most noteworthy due to their influences on the Arg89–Glu374 salt bridge found in HSV-1 TK and cycle 3 TK but absent in HSV-2 TK and cycle 4 TK. These interactions, in turn, influence the mobility of Met128 and Try101 (His101 in the mutants). This finding reinforces the concept that amino acids distant from the active site play important roles in enzyme function, as exemplified by the work of Chessman-Gerber et al.[20] It also highlights the power of DNA family shuffling, which in this case appears to have found at least two different solutions to the same challenge, each incorporating parts of the two different parents as well as novel mutations introduced by the shuffling process.

The altered substrate specificity of the TKs obtained by family shuffling is considerably more pronounced than that reported for TK mutants obtained by cassette mutagenesis[13]. An experimental design feature intended to wean TK away from its preference for thymidine as a substrate, and toward AZT, was the absence of preselection for minimal activity on thymidine, an omission made possible by the use of high-throughput robotic screening of clones that had not been subjected to biological selection for TK activity. Indeed, although cycle 3 and cycle 4 TK retained sufficient activity on thymidine to complement the TK deficiency of *E. coli* KY895, some of the AZT-sensitive clones from the initial round of shuffling did not (data not shown). Furthermore, the kinetic data and structural models for the cycle 3 and cycle 4 TK indicate that the improved activity on AZT came at the expense of activity on thymidine; thus, it is likely that selection for TK activity on thymidine would have limited the change in substrate specificity. Although genetic selections can greatly simplify the library screening process by eliminating inactive clones, our results demonstrate that less restrictive screens can permit the detection of enzymes that have evolved farther away from the wild-type specificity.

The efficacy of AZT–TK therapy, as well as that of several other enzyme-prodrug therapies, depends upon production of the nucleoside analog triphosphate and subsequent incorporation of the analog into DNA. Factors other than AZT phosphorylation that control this efficacy are the activity of AZT monophosphate kinase[21,22], and acceptance of AZT triphosphate by DNA polymerase. One approach to further improving AZT efficacy would be to screen TK libraries for production of AZT diphosphate. An alternative would be to evolve AZT monophosphate kinase activity from thymidylate kinase libraries[10,21–23]. Permissive DNA polymerases such as mammalian polymerase β could be evolved to incorporate AZT triphosphate into DNA, so that the TK, thymidylate kinase, and polymerase genes could be codelivered[24]. Many other suicide gene–prodrug combinations that have been investigated require improvements in enzyme kinetics, and we expect that DNA shuffling could be used to improve them. In addition, efficient delivery of genes remains an impediment to all gene therapy, including anti-HIV efforts, and DNA shuffling offers promise in improving delivery vectors[25].

Family shuffling yields improved progeny that are many mutational steps from the parental genes. We believe that this process yields improvements containing such a large number of amino acid mutations because of the high quality of the proven mutations obtained from natural diversity[6]. These functionally and structurally conservative mutations allow the construction of libraries with high rates of mutations but a low rate of nonfunctional clones. Such libraries sparsely sample a large portion of sequence space and allow rapid functional improvements.

Experimental protocol

Initial library construction. The TK genes from HSV-1 (MacIntyre strain) and HSV-2 (strain G) were isolated from viral DNA (Advanced Biotechnologies, Columbia, MD) by PCR and cloned into a modified pBR322 expression vector. This vector adds a 25-amino acid, N-terminal leader sequence that includes a 6× His-tag sequence for later purification purposes (Fig. 2B). The initial shuffling of the HSV-1 and HSV-2 TK genes was accomplished by recombining randomly DNaseI-digested fragments in a primer-less polymerase reaction as described previously. The shuffled material was cloned back into the same vector described above and transformed into the TK-deficient *E. coli* KY895 (ref. 26) (*E. coli* Genetic Stock Center, New Haven, CT).

AZT screen and cycling. A high-throughput robotic screen identified clones that conferred enhanced AZT sensitivity upon KY895. After transformation, bacteria were plated onto LB media plus ampicillin with no selection for TK activity. The next day, a robotic colony picker (Genetix, Christchurch, UK) was used to transfer colonies to 384-well culture plates, which were incubated overnight. About 10,000 clones per shuffling cycle were spotted from the turbid cultures, again using the robot, onto replicate nylon filters that were then laid over LB agar containing various concentrations of AZT (Sigma, St. Louis, MO). After overnight incubation, visual inspection identified clones that did grow on LB medium but did not grow on LB medium plus AZT. The enhanced AZT-sensitizing phenotype was retested for each clone identified in the primary screen. Confirmed clones were mixed together to provide the genetic material for the next cycle of shuffling and screening. In each successive cycle the screening was made more stringent by reducing the concentration of AZT. To characterize more carefully the best clones identified by screening, they were spread at a low density (about 500 cells per 10-cm plate) on various concentrations of AZT and incubated overnight. Growth inhibition was determined by scraping the bacteria from the plates into saline and measuring the optical density of the solution.

Kinetic measurements. TK proteins were purified from *E. coli* extracts by binding of the 6× His in the leader sequence to nickel columns (Qiagen, Valencia, CA). Better preparations were obtained from *E. coli* TG1 (Pharmacia, Uppsala, Sweden) than from KY895. Kinetics of the HSV-1, HSV-2, cycle 3 TK, and cycle 4 TK enzymes were determined using [*methyl*-^{3}H]thymidine (Amersham, Arlington Heights, IL) and [*methyl*-^{3}H]AZT (Sigma) as substrates in a filter-binding assay essentially as described[13], using conditions such that production of diphosphates was negligible. Data from multiple time points were obtained for multiple substrate concentrations. All kinetic measurements were observed in the linear range, and kinetic constants were calculated from a nonlinear least-squares fit of the initial velocity data to the Michaelis–Menten equation using the computer program HYPERO[27].

Molecular modeling. The three-dimensional model of cycle 4 TK was constructed using the SYBYL 6.3 molecular modeling package (Tripos Associates, St. Louis, MO). The identity of cycle 4 TK to HSV-1 TK is 94%. Thus, the recently determined X-ray structure of HSV-1 TK complexed with dTMP and ADP, solved at 2.75 Å (PDB entry: 1VTK[17]) was used as a template. The starting conformation of AZT has been derived from the crystal structure of azido-thymidine diphosphate bound to nucleoside diphosphate kinase (PDB entry: 1LWX[28]). AZT was modeled in the nucleoside-binding site of TK assuming that the orientation of the pyrimidine ring is equal to that of the base of dTMP. AZT was parameterized for the force field analogs to thymidine, following the procedure described by Perlman and coworkers[29]. Crystallographically determined water molecules, situated either in the active site or in protein cavities, were taken into account. All hydrogen atoms were then added. The wild-type TK and cycle 4 TK complexed with TMP and ADP as well as AZT and ADP were energy minimized using the force field of AMBER 5 (www.amber.ucsf.edu) with the standard parameter set parm96 for the all-atom mode. The ternary complex and crystalline water molecules were centered in a 7.5-Å thick shell of TIP3P water molecules[30]. Any water molecule that had an atom closer than 1.75 Å to any atom of a solute molecule was discarded. A dielectric constant of 1 was used. Nonbonded interactions were calculated within a residue-based cutoff of 10 Å and a secondary cutoff of 12 Å. The structure of the solvated complex was energy minimized by 1,000 steps of steepest descent followed by conjugate gradient until the root-mean-square gradient of the potential energy was less than 0.15 kcal/mol. Finally, the stereochemical quality of the minimized structures was checked using the program PROCHECK[32]. The GRID program[32,33] was used to calculate the most favorable interaction energies between several probes (neutral methyl, sp N with lone pair, sp2 amine NH cation, water, anionic tetrazole N) and HSV-1 TK as well as cycle 4 TK. The GRID default parameters were used. Interaction energies were calculated from a simple nonbonded potential energy function (Lennard-Jones, coulomb, hydrogen bonding) at regularly spaced points of a three-dimensional grid, using a grid resolution of 0.5 Å (ref. 32). Energy levels indicating the most favorable binding sites were displayed as hypersurfaces on an SGI Indigo2 Extreme graphic workstation (Silicon Graphics, Mountain View, CA) using the SYBYL 6.3 molecular modeling package.

Acknowledgments

Thanks to Sun-Ai Raillard, Glenn Dawes, and Steve delCardayre for technical assistance, to the Computational Center of the ETH Zürich for computer time, and to Claus Krebber, Larry Loeb, Frances Arnold, and Rolf Zinkernagel for comments on the manuscript.

1. Stemmer, W.P.C. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91,** 10747–10751 (1994).
2. Stemmer, W.P.C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370,** 389–391 (1994).
3. Zhang, J.-H., Dawes, G. & Stemmer, W.P.C. Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci. USA* **94,** 4504–4509 (1997).
4. Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P.C. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.* **14,** 315–319 (1996).
5. Buchholz, F., Angrand, P.-O. & Stewart, A.F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **16,** 657–662 (1998).
6. Crameri, A., Raillard, S.-A., Bermudez, E. & Stemmer, W.P.C. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391,** 288–291 (1998).
7. Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T. & Furukawa, K. Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* **16,** 663–666 (1998).
8. Hirsch, M.S., Kaplan, J.C. & D'Aquila, R.T. in *Fields virology* (eds. Fields, B.N., Knipe, D.M. & Howley, P.M.) 431–466 (Lippincott-Raven, Philadelphia, 1996).
9. Martin, L.-A. & Lemoine, N.R. Direct cell killing by suicide genes. *Cancer Metastasis Rev.* **15,** 301–316 (1996).
10. Guettari, N., Loubiere, L., Brisson, E. & Klatzmann, D. Use of herpes simplex virus thymidine kinase to improve the antiviral activity of zidovudine. *Virology* **235,** 398–405 (1997).
11. Drake, R.R. et al. Metabolism and activities of 3′-azido-2′,3′-dideoxythymidine and 2′,3′-didehydro-2′,3′-dideoxythymidine in herpesvirus thymidine kinase transduced T-lymphocytes. *Antiviral Res.* **35,** 177–185 (1997).
12. Gentry, G.A. Viral thymidine kinases and their relatives. *Pharmacol. Ther.* **54,** 319–355 (1992).
13. Munir, K.M., French, D.C. & Loeb, L.A. Thymidine kinase mutants obtained by random sequence selection. *Proc. Natl. Acad. Sci. USA* **90,** 4012–4016 (1993).
14. Black, M.E. & Loeb, L.A. Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants in vitro. *Biochemistry* **32,** 11618–11626 (1993).
15. Black, M.E., Newcomb, T.G., Wilson, H.-M.P. & Loeb, L.A. Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl. Acad. Sci. USA* **93,** 3525–3529 (1996).
16. Moore, J.C., Jin, H.-M., Kuchner, O. & Arnold, F.H. Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* **273,** 336–347 (1997).
17. Wild, K., Bohner, T., Folkers, G. & Schulz, G.E. The structures of thymidine kinase from *Herpes simplex* virus type 1 in complex with substrates and a substrate analog. *Protein Sci.* **6,** 2097–2106 (1997).
18. Brown, D.G. et al. 1995. Crystal structures of the thymidine kinase from herpes simplex virus type-1 in complex with deoxythymidine and ganciclovir. *Nat. Struct. Biol.* **2,** 876–881 (1995).
19. Champness, J.N. et al. Exploring the active site of herpes simplex virus type-1 thymidine kinase by x-ray crystallography of complexes with acyclovir and other ligands. *Struct. Funct. Genet.* **32,** 350–361 (1998).
20. Kussmann-Gerber, S., Kuonen, O., Folkers, G., Pilger, B.D. & Scapozza, L. Drug resistance of herpes simplex virus type 1: structural considerations on the molecular level of the thymidine kinase. *Eur. J. Biochem.* **255,** 472–481 (1998).
21. Lavie, A. et al. Structure of thymidylate kinase reveals the cause behind the limiting step in AZT activation. *Nat. Struct. Biol.* **4,** 601–604 (1997).
22. Lavie, A. et al. The bottleneck in AZT activation. *Nat. Med.* **3,** 922–924 (1997).
23. Balzarini, J. et al. Improving AZT efficacy. *Nat. Med.* **4,** 132 (1998).
24. Bouayadi, K. et al. Overexpression of DNA polymerase b sensitizes mammalian cells to 2′,3′-didexoycytidine and 3′-azido′-3′-deoxythymidine. *Cancer Res.* **57,** 110–116 (1997).
25. Patten, P.A., Howard, R.J. & Stemmer, W.P.C. Applications of DNA shuffling to pharmaceuticals and vaccines. *Curr. Opin. Biotechnol.* **8,** 724–733 (1997).
26. Igarashi, K., Hiraga, S. & Yura, T. A deoxythymidine kinase deficient mutant of Escherichia coli. II. Mapping and transduction studies with phage phi 80. *Genetics* **57,** 643–654 (1967).
27. Cleland, W.W. Statistical analysis of enzyme data. *Methods Enzymol.* **63,** 103–138 (1979).
28. Xu, Y. et al. X-ray analysis of azido-thymidine diphosphate binding to nucleoside diphosphate kinase. *Proc. Natl. Acad. Sci. USA* **94,** 7162–7165 (1997).
29. Perlman, D.A. et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.* **91,** 1–41 (1995).
30. Jorgensen, W.L., Chrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
31. Laskowski, R.A., MacArthur, M.W., Moos, D.S. & Thornton, J.M.J. PROCHECK: a program to check the stereochemical quality of protein structures. *Appl. Cryst.* **26,** 283–291 (1993).
32. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28,** 849–857 (1985).
33. Wade, R.C. & Goodford, P.J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **36,** 140–156 (1993).

Protein evolution by molecular breeding

Jeremy Minshull* and Willem PC Stemmer†

Natural evolution has guided the development of 'molecular breeding' processes used in the laboratory for the rapid modification of subgenomic sequences including single genes. The most significant recent development has been the *in vitro* permutation of natural diversity. Homologous recombination of multiple related sequences produced high-quality libraries of chimeric sequences encoding proteins with functions that differ dramatically from any of the parents. Increasingly powerful screening methods are also being developed, allowing these libraries to be screened for novel biocatalysts.

Addresses
Maxygen Incorporated, 515 Galveston Drive, Redwood City, CA 94063, USA
*e-mail: jeremy_minshull@maxygen.com
†e-mail: pim_stemmer@maxygen.com

Current Opinion in Chemical Biology 1999, 3:284–290

http://biomednet.com/elecref/1367593100300284

 ISSN 1367-5931

Introduction

Enzymes are used in a wide variety of applications including food and feed processing, laundry detergents, chemicals production, paper bleaching and pharmaceutical manufacturing. The benefits of using enzymes as catalysts are that reactions can occur at moderate temperatures, toxic solvents or reactants can often be eliminated, and reactions are usually stereospecific, which is of particular benefit in the synthesis of pharmaceuticals and fine chemicals. The specificity of enzymes also obviates the need for protecting and deprotecting reactive groups, which is a source of considerable yield loss in organic syntheses.

Although three billion years of evolution have produced a wealth of protein catalysts, they are generally not optimal for a particular industrial application. While it is possible to screen enzymes from extremophiles for activity under the appropriate process reaction conditions [1,2], natural selection has selected enzymes to function in the complex mixtures of molecules within cells rather than in bioreactors. Obtaining the desired combinations of properties therefore generally requires further protein optimization.

Structural information has been used with some success to improve enzyme function [3–5]. As a general method, however, structure-based methods require time and equipment in order to generate and process very large amounts of information.

An alternative strategy to making defined changes on the basis of structural understanding is to harness the Darwinian power of recursive cycles of mutation and selection. By using directed evolution, protein engineers attempt to mimic the natural processes by which protein variants arise and are tested for 'fitness' within living systems. In this review, we will focus on the underlying rationale behind and recent advances in directed evolution, both in the methods used to generate protein variants, and in the screening strategies used to identify variants of interest.

DNA shuffling

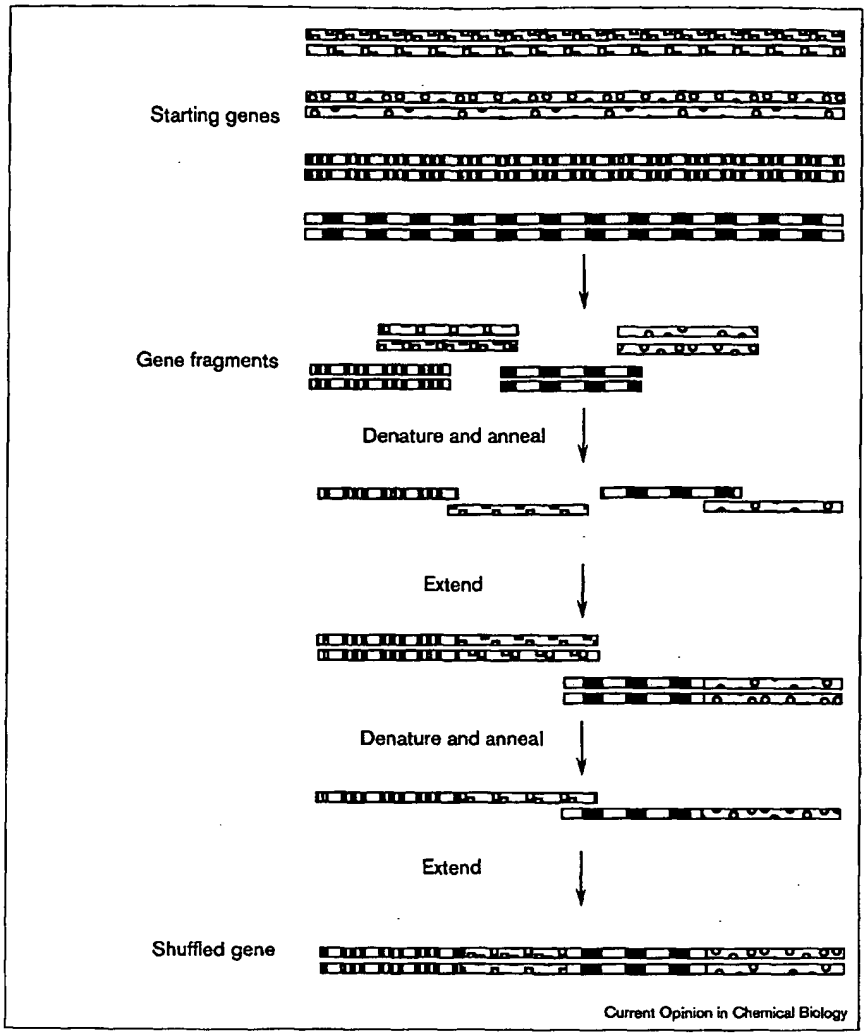
Directed evolution effectively performs the complex computations required to determine the effects of changes in sequence on catalytic function. In addition to the active-site geometry, the impact of sequence changes on protein expression, stability and folding, and interactions with other host proteins and small molecules are all simultaneously considered simply by directly measuring the activity of the mutant enzymes or metabolic pathways.

The best evolutionary strategies are likely to be those that most closely mimic natural ones: in three billion years, not only have individual genes evolved, but the evolutionary process itself has been optimized [6]. Those algorithms that are best at searching through the possible combinations of nucleotides for sequences with biological function have been preserved along with the sequences whose evolution they have facilitated. Recombination is such a mechanism, found universally in biological systems. Genetic algorithms and other computer simulations of simple evolving systems that incorporate the ability to recombine information are more powerful and evolve more rapidly than those which do not [6–9].

Incorporation of recombination into a method for directed evolution of single genes (known as 'DNA shuffling' or 'molecular breeding') was developed recently [10]. In this method, a population of mutant genes (rather than just one) are selected on the basis of their containing beneficial mutations, thus making them appropriate as parents for the next cycle. The genes are randomly fragmented, then reassembled by recombination with each other. The process is shown schematically in Figure 1. As well as accelerating the *in vitro* evolutionary process [10–12], the shuffling reaction is extremely flexible: many different pieces of genetic information may be included if they are available (see Figure 2; [13]). For example, Liu *et al.* [14] included degenerate oligonucleotides in their shuffling reaction in order to randomize amino acids believed, through structural studies, to be important for the substrate specificity of a tRNA synthase. Interestingly, only one of the five targeted residues was mutated in the enzyme showing highest activity against the new substrate.

Figure 1

In vitro recombination by DNA shuffling. Genes are fragmented and then reassembled by a reaction in which homologous fragments act as primers for each other.



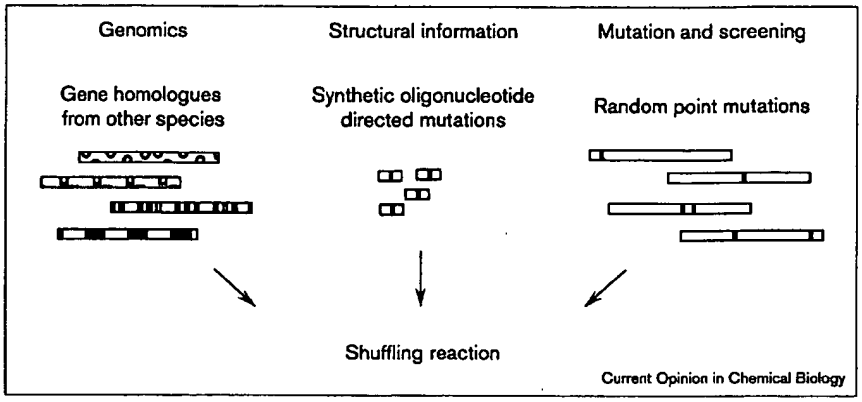
Many examples of successful directed evolution using DNA shuffling have been reviewed recently [15•,16•]. Last year, several additional formats were described for the *in vitro* [17,18] or *in vivo* [19] shuffling of genes. While these methods have not been thoroughly compared, they rely on the same underlying principle that the most efficient way to explore all of the possible combinations and permutations of sequences (i.e. sequence space) is by recombination of active variants.

Screening and selection

Natural evolution measures the fitness of variants by their ability to survive. In some cases, there are genetic selections that can be employed to make a cell's growth dependent on a particular improved function. Schellenberger's group [20•] recently selected for increased subtilisin production by making a target protein the sole source of nitrogen, performing the growth in hollow fibres to prevent cross-feeding. As an artificial selection system, phage display has been used to identify proteins that bind specific ligands. Catalytic proteins displayed on phage have also been selected, either by making infectivity dependent on formation of a covalent intermediate [21••], or by requiring enzyme activity to release the phage from a solid matrix [22•]. Both of these methods only require a single catalytic event, so are unsuitable for quantitative measurements.

Directed evolution has been used to enhance lipase enantioselectivity. Lipases accept a wide variety of non-natural esters, so lipases that are able to discriminate between stereoisomers allow the production of optically pure compounds useful in pharmaceutical and fine chemical manufacture. One group used a microtitre-based absorbance assay in which the esterase activity of lipase variants was measured against the *R* and *S* forms of *p*-nitrophenyl 2-methyldecanoate. Four cycles, testing 1,000 lipase mutants per cycle, increased the enantioselectivity from 2% enantiomeric excess (ee) to 81% ee in favor of the *S* configuration [23]. A second group evolved an

Figure 2



The shuffling reaction is extremely flexible. Positive variants resulting from random mutation and selection can be recombined with sequence information obtained computationally. Genomics allow the inclusion of related genes from other species and structural information can be used to design synthetic oligonucleotides for making specific changes or to randomize targeted regions of a protein.

enzyme to hydrolyze an ester for production of an intermediate in epithilone synthesis. The initial screen for this enzyme was performed by including both the enzyme substrate and a pH indicator in agar plates. Bacterial colonies expressing an enzyme able to hydrolyze the ester were identified by a change in the colour of the indicator, since acid is released when esters are hydrolyzed. Colonies selected by this screen were then picked and tested for their biotransformation activity and stereoselectivity by measuring the optical rotation of the products [24•]. While individual screens will always depend on the reactions being catalyzed, this strategy of tiered screening in which a primary, relatively inaccurate assay is used to select a small number of clones that are then subjected to more detailed analysis (see Figure 3) is an extremely powerful general technique.

It is also possible to perform an entire selection *in vitro*. As an example, a library of genes was transcribed and translated in compartments formed in a water/oil emulsion. Active DNA methyltransferase *Hae*III enzymes methylated the genes that encoded them, thereby protecting the DNA from subsequent *Hae*III digestion [25••]. By using such a system, cloning or transformation of the library is not required, so much larger libraries can be screened. Further advances such as coupled reactions leading to gene modification and sorting of intact compartments based on fluorescence would help make *in vitro* enzyme production and testing a very powerful methodology.

Using natural diversity

In addition to developing screening strategies that allow greater numbers of mutants to be screened, directed evolution can be optimized by building protein libraries that contain the maximum number of active (and different) members. Until this year, single genes were used as starting points for DNA shuffling and variants, arising by point mutation, were very similar in sequence to the parent gene. Another approach uses principles similar to those of the mammalian immune system. Antibodies capable of binding essentially any epitope with nanomolar association constants are generated by recombinination between a few thousand sequences, followed by 'affinity maturation' by point mutation [26]. Enzyme catalysis results from binding to and stabilizing the relevant transition-state analogue [27], so it should be possible to harness such a system to produce enzymes [28]. Antibodies have evolved as rigid binding molecules, however, and catalytic antibodies are selected solely by their abilities to bind transition-state analogues rather than other enzymatically essential functions such as substrate binding and product release. They are thus generally much less active as catalysts than proteins that have evolved as enzymes.

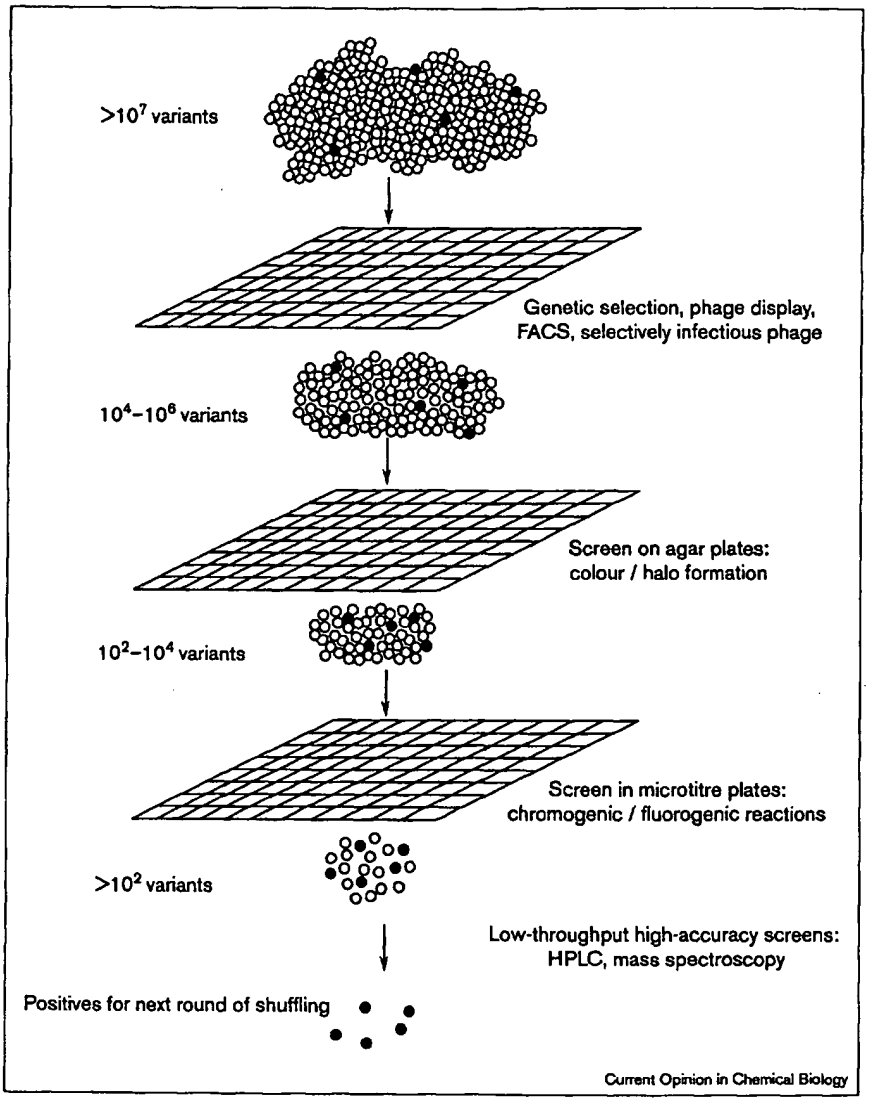
Instead of trying to turn antibodies into catalysts, DNA shuffling can be used to mimic the immune system's incredibly powerful diversity-generating process, by recombining genes with one another. In the first example of 'DNA family shuffling', four different β-lactamase genes were shuffled together to produce a chimera with 270-fold greater resistance to moxalactam than the best parental enzyme [29•]. The chimeric enzyme produced in this experiment differed from each parent by at least 100 amino acids (Figure 4), yet was still a fully functional cephalosporinase. Like antibody 'diversity' regions, sequences that occur in naturally existing enzymes have already been tested for their ability to function within the context of the protein's overall structure. Recombining natural blocks of sequence with each other allows a broad region of functional sequence space to be sampled sparsely.

Protein chimeras may differ dramatically from all their parents

Where an active site lies at the interface between folding subdomains, exchanging these subdomains will alter the shape of the active site. For example, swapping domains between coagulation factor X and trypsin produced a serine protease with broadened substrate selectivity [30•]. The activities of chimeric enzymes are often not predictable simply by comparing those of the parent enzymes,

Figure 3

Tiered screening. Variants are tested by a series of assays that are successively more accurate and more time- and labour-intensive. It is important to ensure that the higher capacity assays correlate well with the desired final activity. FACS, fluorescence-activated cell sorting; HPLC, high-pressure liquid chromatography.

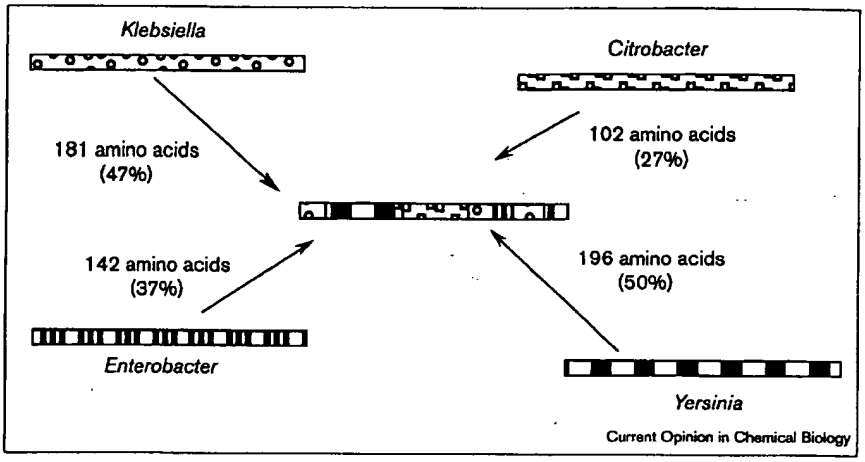


as was found for chimeras between two human blood group glycosyl transferases that were shown to be functionally interconvertible by changing only four amino acids. Parental enzyme A transfers *N*-acetylgalactosamine to a disaccharide acceptor, whereas enzyme B transfers galactose. Replacement of Arg176 in enzyme A with the Gly176 of enzyme B resulted not in increases in B-like activities, but in a fourfold higher k_{cat}/K_M for the enzyme A substrate (i.e. *N*-acetylgalactosamine) [31••].

Altered substrate specificities have also been produced by random recombination of sequences followed by screening. Biphenyl dioxygenases initiate the degradation of polychlorinated biphenyls, and their congener substrate specificities are determined by the large terminal subunit [32]. DNA shuffling of two such dioxygenases produced chimeras with a different substrate range from either parent, enhanced degradation of biphenyl compounds and even novel oxygenation activity for single aromatic hydrocarbons [33,34••].

Random chimeras have also been made *in vivo* between two staphylococcal lipases with differing chain-length selectivities and phospholipase activities. Novel enzymes were found that possessed both combinations of and absolute levels of these activities that differed from both parents in ways that were often surprising [35•]. For example, one chimera in which a block comprising 20% of the enzyme with no chain-length selectivity was incorporated into the enzyme with a strong preference for short-chain fatty acids unexpectedly resulted in an enzyme with twofold increased activity (relative to the best parent) against the long-chain ester *p*-nitrophenyl palmitate.

Figure 4



Mutational distances of chimeric β-lactamase with 270-fold improved moxolactamase activity from its four parents. Distances from each parent are given in number of amino acids, and in the percentage of residues that this represents. The chimera differs by 102 amino acids, that is 27% of positions, from its closest parent (the *Citrobacter* enzyme). It would not be possible to make 102 random changes without inactivating the enzyme. Thus recombination of natural diversity allows functional sequence space to be sampled much more broadly and sparsely than sequential point mutations from a single starting sequence.

Recursive cycles of shuffling using multiple parents has been performed by Christians *et al* [36••]. By recombining two Herpes Simplex Virus thymidine kinase genes and robotically screening for variants that were better able to phosphorylate the therapeutic nucleotide analogue AZT, the concentration of AZT required to inhibit cell growth was reduced 32-fold relative to that required with the best parent. The resulting enzyme was a chimera that had undergone ten cross-over events between the two parental genes, and had also accumulated five point mutations, leading to a protein differing by 22 amino acids from the closest parent. The process of recombination between different but functional parents to make large changes in sequence, coupled with point mutagenesis to fine-tune the activity of the protein, is highly analogous to the process of antibody generation and maturation.

Directed searches for novel protein activities

Although it is possible to modify the physical properties of an enzyme, such as thermostability or activity in organic solvent, by screening for sequential improvements in these properties [37–39], modification of one property by single point mutations can often compromise another desired characteristic [40•]. From the results discussed above, we would predict that by recombining sequences found in nature, it should be possible to discover enzymes possessing all combinations of properties of the individual parents, as well as improvements over any of the parents.

The classification of enzymes into superfamilies that appear to be related by a common chemical strategy for stabilizing the transition state for the formation of a reactive intermediate suggests a mechanism by which nature may evolve novel catalytic functions [41]. Is it possible to make such changes in the laboratory? It may not be possible to make a graded change from one reaction to another. By making structural comparisons between an oleate desaturase and an oleate hydroxylase, Broun *et al.* [42••] have shown that four amino acid changes in the desaturase can convert it to a hydroxylase and changing six residues in the hydroxylase result in desaturase activity. Making these changes by sequential point mutagenesis would not be possible because the single or double mutants do not possess intermediate activities. The exchange of blocks of amino acids made possible by family shuffling, however, offers a possible route to completely novel substrate specificities. Enzyme libraries constructed from relatively small families of homologous genes are likely to contain not only a range of substrate specificities, but also a variety of physical properties and even new catalytic activities. These libraries can then serve as sources of diversity themselves, providing the starting points for further directed evolution in many different directions.

Conclusions

By copying the natural mechanisms by which even existing diversity can be recombined, DNA shuffling can be used to generate high-quality libraries of novel proteins. Chimeras between naturally occurring enzymes that differ by only a few amino acids often possess activities that are significantly different from their parents. By screening these libraries using innovative high-throughput assay techniques, it is possible to identify enzymes with new catalytic functions and physical properties.

Acknowledgements

We thank SunAi Raillard, Claus Krebber, Andreas Crameri, Abby Dernburg and Matt Tobin for stimulating discussions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Robertson DE, Mathur EJ, Swanson RV, Marrs BL, Short JM: **The discovery of new biocatalysts from microbial diversity.** *Soc Indust Microbiol News* 1996, 46:3-8.

2. Madigan MT, Marrs BL: **Extremophiles.** *Sci Am* 1997, **276**:82-87.

3. Rubingh DN: **Protein engineering from a bioindustrial point of view.** *Curr Opin Biotechnol* 1997, **8**:417-422.

4. Hellinga HW: **Rational protein design: combining theory and experiment.** *Proc Natl Acad Sci USA* 1997, **94**:10015-10017.

5. Shao Z, Arnold FH: **Engineering new functions and altering existing functions.** *Curr Opin Struct Biol* 1996, **6**:513-518.

6. Kelly K: *Out of Control: The Rise Of Neo-Biological Civilization.* Menlo Park, California: Addison–Wesley; 1994.

7. Holland JH: *Adaptation In Natural And Artificial Systems.* Ann Arbor, MI: University of Michigan Press; 1975.

8. Forrest S: **Genetic algorithms: principles of natural selection applied to computation.** *Science* 1993, **261**:872-878.

9. Gibson JM: **Simulated evolution and artificial selection.** *Biosystems* 1989, **23**:219-229.

10. Stemmer WP: **Rapid evolution of a protein *in vitro* by DNA shuffling.** *Nature* 1994, **370**:389-391.

11. Stemmer WPC: **Searching sequence space.** *Biotechnology* 1995, **13**:549-553.

12. Moore JC, Jin HM, Kuchner O, Arnold FH: **Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences.** *J Mol Biol* 1997, **272**:336-347.

13. Stemmer WP: **DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution.** *Proc Natl Acad Sci USA* 1994, **91**:10747-10751.

14. Liu DR, Magliery TJ, Pastrnak M, Schultz PG: **Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*.** *Proc Natl Acad Sci USA* 1997, **94**:10092-10097.

15. • Giver L, Arnold FH: **Combinatorial protein design by *in vitro* recombination.** *Curr Opin Chem Biol* 1998, **2**:335-338.
This is a good summary of several examples of protein evolution by recombination-based methods.

16. • Harayama S: **Artificial evolution by DNA shuffling.** *Trends Biotechnol* 1998, **16**:76-82.
This is a review of DNA shuffling examples.

17. Shao Z, Zhao H, Giver L, Arnold FH: **Random-priming *in vitro* recombination: an effective tool for directed evolution.** *Nucleic Acids Res* 1998, **26**:681-683.

18. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH: **Molecular evolution by staggered extension process (StEP) *in vitro* recombination.** *Nat Biotechnol* 1998, **16**:258-261.

19. Zhang Y, Buchholz F, Muyrers JPP, Stewart AF: **A new logic for DNA engineering using recombination in *Escherichia coli*.** *Nat Genet* 1998, **20**:123-128.

20. • Naki D, Paech C, Granshaw G, Schellenberger V: **Selection of a subtilisin-hyperproducing *Bacillus* in a highly structured environment.** *Appl Microbiol Biotechnol* 1998, **49**:290-294.
This is a selection designed to link increased protease activity to faster growth rate. The authors identified a mutant with increased expression, but is also a more active protease.

21. •• Gao C, Lin CH, Lo CHL, Mao S, Wirsching P, Lerner RA, Janda KD: **Making chemistry selectable by linking it to infectivity.** *Proc Natl Acad Sci USA* 1997, **94**:11777-11782.
Linking genotype to phenotype is essential for evolution. Here, phage infectivity requires a catalytically active protein to be fused to the gene III protein. The authors were only able to detect single turnover events, but the technique represents a potential way to screen large libraries for novel catalytic function.

22. • Pedersen H, Holder S, Sutherlin DP, Schwitter U, King DS, Schultz PG: **A method for directed evolution and functional cloning of enzymes.** *Proc Natl Acad Sci USA* 1998, **95**:10523-10528.
The authors describe another way of using phage display to identify catalysts. They showed that binding to or release from a physical support is dependent on a single turnover event.

23. Reetz MT, Zonta A, Schimossek K, Liebeton K, Jaeger K-E: **Creation of enantioselective biocatalysts for organic chemistry by *in vitro* evolution.** *Agnew Chem Int Ed Engl* 1997, **36**:2830-2832.

24. • Bornscheuer UT, Altenbuchner J, Meyer HH: **Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epithilones.** *Biotechnol Bioeng* 1998, **58**:554-559.
A tiered screen is described, and directed evolution of enantioselectivity is presented.

25. •• Tawfik DS, Griffiths AD: **Man-made cell-like compartments for molecular evolution.** *Nat Biotechnol* 1998, **16**:652-656.
An elegant *in vitro* transcription, translation and selection system is described in which cloning and transformation are things of the past. Some major limitations still need to be overcome to allow general applicability, but one day all enzymes will be evolved like this.

26. Tonegawa S: **Somatic generation of antibody diversity.** *Nature* 1983, **302**:575-581.

27. Pauling L: **Molecular architecture and biological reactions.** *Chem Eng News* 1946, **24**:1375-1377.

28. Jencks WP: *Catalysis In Chemistry And Enzymology.* New York: McGraw Hill; 1969.

29. • Crameri A, Raillard S-A, Bermudez E, Stemmer WPC: **DNA shuffling of a family of genes from diverse species accelerates directed evolution.** *Nature* 1998, **391**:288-291.
Here, generation of an improved variant enzyme that is very different from it's parents is described. The authors demonstrate that chimeragenesis is a good way to sample broad areas of sequence space sparsely.

30. • Hopfner K-P, Kopetzki E, Kreße G-B, Bode W, Huber W, Engh RA: **New enzyme lineages by subdomain shuffling.** *Proc Natl Acad Sci USA* 1998, **95**:9813A.
A deliberate attempt is made to change the shape of the catalytic cleft by making a chimera with folding subdomains from different proteases. Novel but unpredicatable properties result.

31. •• Seto NOL, Palcic MM, Compston CA, Li H, Bundle DR, Narang SA: **Sequential interchange of four amino acids from blood group B to blood group A glycosyltransferase boosts catalytic activity and progressively modifies substrate recognition in human recombinant enzymes.** *J Biol Chem* 1997, **272**:14133-14138.
Exchange of very limited natural diversity between two functionally distinct enzymes produces chimeras with unexpected activities is described.

32. Erickson BD, Mondello FJ: **Enhanced biodegradation of polychlorinated biphenyls after site-directed mutagenesis of a biphenyl dioxygenase gene.** *Appl Environ Microbiol* 1993, **59**:3858-3862.

33. Kimura N, Nishi A, Goto M, Furukawa K: **Functional analyses of a variety of chimeric dioxygenases constructed from two biphenyl dioxygenases that are similar structurally but different functionally.** *J Bacteriol* 1997, **179**:3936-3943.

34. •• Kumamaru T, Suenaga H, Mitsuoka M, Watanabe T, Furukawa K: **Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase.** *Nat Biotechnol* 1998, **16**:663-666.
DNA shuffling of two highly homologous large subunits of multisubunit enzymes is described. Chimeras possess novel substrate preferences and are still able to interact with the other components of the enzyme system.

35. • van Kampen MD, Dekker N, Egmond MR, Verheij HM: **Substrate specificity of *Staphylococcus* hylcus lipase and *Staphylococcus aureus* lipase as studied by *in vivo* chimeragenesis.** *Biochemistry* 1998, **37**:3459-3466.
The authors describe the use of chimeras to crudely correlate primary sequence with enzyme substrate specificity. The main conclusion drawn is that chimeric proteins can exceed the sum of their parts. Quite right!

36. •• Christians FC, Scapozza L, Crameri A, Folkers G, Stemmer WPC: **Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling.** *Nat Biotechnol* 1999, **17**:259-264.
Chimeragenesis, multiple rounds of DNA shuffling and a robotic screen for improved function are described. Also, a nice test of how many mutants one should take from one round to the next is discussed.

37. Zhao H, Arnold FH: **Directed evolution converts subtilisin E into a functional equivalent of thermitase.** *Protein Eng* 1999, **12**:47-53.

38. Akanuma S, Yamagishi A, Tanaka N, Oshima T: **Serial increase in the thermal stability of 3-isopropylmalate dehydrogenase from Bacillus subtilis by experimental evolution.** *Protein Sci* 1998, **7**:698-705.

39. Buchholz F, Angrand P-O, Stewart AF: **Improved proerties of FLP recombinase evolved by cycling mutagenesis.** *Nat Biotechnol* 1998, **16**:657-662.

40. • Patkar S, Vind J, Kelstrup E, Christensen MW, Svendsen A, Borch K, Kirk O: **Effect of mutations in *Candida antarctica* B lipase.** *Chem Phys Lipids* 1998, **93**:95-101.
The authors juggle with thermostability, enantioselectivity, specific activity and oxidation stability of a lipase. They conclude that changes that improve one property degrade another.

41. Gerlt JA, Babbitt PC: **Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis.** *Curr Opin Chem Biol* 1998, **2**:607-612.

42. •• Broun P, Shanklin J, Whittle E, Somerville C: **Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids.** *Science* 1998, **282**:1315-1317.
Small numbers of amino acid substitutions change not only the substrate specificity of an enzyme, but also the reaction catalysed. Watch in awe as a desaturase becomes a hydroxylase.

APPENDIX F

Attorney Docket No. 035718/237005 PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re:	Abad *et al.*	Confirmation No.:	5409
Appl. No.:	10/032,717	Group Art Unit:	1638
Filed:	10/23/2001	Examiner:	A.R. Kubelik
For:	GENES ENCODING NOVEL BACILLUS THURINGIENSIS PROTEINS WITH PESTICIDAL ACTIVITY AGAINST COLEOPTERANS		

January 18, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RULE 132 DECLARATION
of
Carl Simmons

Sir:

I, Carl Simmons, do hereby declare and say as follows:

1. I am skilled in the art of the field of the invention. I have earned a Bachelor of Science in Genetics and Cell Biology from the University of Minnesota in 1985. I earned Masters (1987) and Ph.D. (1991) degrees in Genetics from the University of California, Davis, with a thesis on molecular characterization of rice and cereal genes. I served as a post-doctoral associate at Yale University in the Biochemistry and Biophysics Department from 1991-1995, where I studied molecular genetics and physiology in *Arabidopsis*. I joined Pioneer Hi-Bred International, Inc. in 1995 and worked in the Disease Resistance group until 1998, with an emphasis on maize gene discovery and functional evaluation. I joined the DuPont-Pioneer Bioinformatics group in 1998, where I have been actively researching a wide variety of projects in the area of genomics, gene expression, and gene and protein structure/function.. I have authored dozens of scientific papers and patents in the area of plant molecular biology, gene identification, and bioinformatics.

2. I collaborated with Tom McNeill to perform of the analysis described in the Response in the above case to the Office Action of 10/21/2004 and believe that it correctly represents the difference between the SEQ ID NO:1 of the present claims in the above application and the nucleotide sequences that encode the cited Michaels protein. Accordingly, the analysis shows that nucleotide sequences having 90% and higher sequence identity to the SEQ ID NO:1 disclosed in the above-referenced application could not encode the protein having the amino acid sequence set forth in SEQ ID NO:4 of the Michaels reference, U.S. Patent 5,554,534.

3. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 1/20/05 By: _____

Carl Simmons

APPENDIX G

Attorney Docket No. 035718/237005
PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re:	Abad *et al.*	Confirmation No.:	5409
Appl. No.:	10/032,717	Group Art Unit:	1638
Filed:	10/23/2001	Examiner:	A.R. Kubelik

For: GENES ENCODING NOVEL BACILLUS THURINGIENSIS PROTEINS WITH PESTICIDAL ACTIVITY AGAINST COLEOPTERANS

January 18, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RULE 132 DECLARATION
of
Thomas Z. McNeill

Sir:

I, Thomas Z. McNeill, do hereby declare and say as follows:

1. I am skilled in the art of the field of the invention. I have earned a Bachelor of Science in Biology from Trinity University in 1993. I earned a Masters (1999) and Ph.D. (2004) degrees in Biology from the University of Houston in Houston, Texas, with a dissertation on computational methods for genome annotation and have authored several papers in the area. I joined Pioneer Hi-Bred International, Inc. in 2003 in the Bioinformatics and Exploratory Research Group with an emphasis in scientific software development.

2. I collaborated with Carl Simmons to perform of the analysis described in the Response in the above case to the Office Action of 10/21/2004 and believe that it correctly represents the difference between the SEQ ID NO:1 of the present claims in the above application and the nucleotide sequences that encode the cited Michaels protein. Accordingly, the analysis shows that

nucleotide sequences having 90% and higher sequence identity to the SEQ ID NO:1 disclosed in the above-referenced application could not encode the protein having the amino acid sequence set forth in SEQ ID NO:4 of the Michaels reference, U.S. Patent 5,554,534.

3. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 1/20/05

By: _____
Thomas Z. McNeill

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.